

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2000, 10:30:20 ; Search time 3022.29 Seconds  
(without alignments)  
2182.479 Million cell updates/sec

Title: US-09-215-435-66  
Perfect score: 1747  
Sequence: 1 gccaccacaaatgttcctt.....tggtcaaaaaaaaaaaaaa 1747

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues 9077268  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: gb\_est1: \*  
21: gb\_est2: \*  
22: gb\_est3: \*  
23: gb\_est4: \*  
24: gb\_est5: \*  
25: gb\_est6: \*  
26: gb\_est7: \*  
27: gb\_est8: \*  
28: gb\_est9: \*  
29: gb\_est10: \*  
30: gb\_est11: \*  
31: gb\_est12: \*  
32: gb\_est13: \*  
33: gb\_est14: \*  
34: gb\_est15: \*  
35: gb\_est16: \*  
36: gb\_est17: \*  
37: gb\_est18: \*  
38: gb\_est19: \*  
39: gb\_est20: \*  
40: gb\_est21: \*  
41: gb\_est22: \*  
42: gb\_est23: \*  
43: gb\_est24: \*  
44: gb\_est25: \*

45: gb\_est26: \*  
46: gb\_est27: \*  
47: gb\_est28: \*  
48: gb\_est29: \*  
49: gb\_est30: \*  
50: gb\_est31: \*  
51: gb\_est32: \*  
52: em\_est20: \*  
53: em\_est21: \*  
54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*  
59: gb\_est33: \*  
60: gb\_est34: \*  
61: gb\_est35: \*  
62: gb\_est36: \*  
63: gb\_est37: \*  
64: gb\_est38: \*  
65: em\_est27: \*  
66: em\_est28: \*  
67: em\_est29: \*  
68: em\_est30: \*  
69: gb\_est39: \*  
70: gb\_est40: \*  
71: gb\_est41: \*  
72: gb\_est42: \*  
73: gb\_est43: \*  
74: gb\_est44: \*  
75: em\_est31: \*  
76: em\_est32: \*  
77: em\_est33: \*  
78: em\_est34: \*  
79: gb\_gss1: \*  
80: gb\_gss2: \*  
81: gb\_gss3: \*  
82: gb\_gss4: \*  
83: em\_gss1: \*  
84: em\_gss2: \*  
85: em\_gss3: \*  
86: em\_gss4: \*  
87: gb\_gss5: \*  
88: gb\_gss6: \*  
89: gb\_gss7: \*  
90: gb\_gss8: \*  
91: gb\_gss9: \*  
92: em\_gss5: \*  
93: em\_gss6: \*  
94: em\_gss7: \*  
95: em\_gss8: \*  
96: em\_gss9: \*  
97: em\_gss10: \*  
98: em\_gss11: \*  
99: gb\_gss10: \*  
100: gb\_gss11: \*  
101: em\_gss12: \*  
102: gb\_gss12: \*  
103: gb\_gss13: \*  
104: gb\_gss14: \*  
105: gb\_gss15: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	973.8	55.7	1114	51	AF150387
2	493.2	28.2	525	41	AI005033

AF150387 AF150387  
AI005033 ou91b12.x

```

c 3 434.8 24.9 464 37 AA701598 z134h07.s
c 4 434.8 24.9 464 42 A1127789 qc32c02.x
c 5 417.8 23.9 443 42 A1127822 qc36f02.x
c 6 414.8 23.7 444 42 A1095556 qb19f04.x
c 7 397.2 22.7 428 42 A1094243 qa43c11.s
c 8 394.4 22.6 447 64 AW015855 UI-H-B10-
c 9 386.8 22.1 427 37 AA677847 z113c06.s
c 10 379.2 21.7 448 22 R55298 yj77e01.s1
c 11 377.8 21.6 411 44 A1248727 qb72d06.x
c 12 375.8 21.5 425 61 A1857994 wj69a09.x
c 13 368.2 21.1 450 29 A136368 zx93d05.r
c 14 365.4 20.9 398 41 A1040388 ox16d04.x
c 15 343.8 19.7 420 22 H15682 ym27g11.r1
c 16 339.8 19.5 369 42 A1081812 ox77f03.x
c 17 337.6 19.3 402 22 R55398 yj77e01.r1
c 18 310.4 17.8 341 62 A1874047 ym47f12.x
c 19 299.4 17.1 327 37 AA705405 zj90d03.s
c 20 296.4 17.0 310 21 F13007 HSC3HD121 n
c 21 289 16.5 301 37 AA723830 ah63b03.s
c 22 276 15.8 399 30 AA207232 zq61f07.r
c 23 275.8 15.8 455 22 H14538 y125d10.r1
c 24 274.6 15.7 357 40 AA982239 ua49c03.r
c 25 274.4 15.7 315 50 A1669838 tu31h11.x
c 26 272.6 15.6 434 29 A136283 zK93d05.s
c 27 272.2 15.6 368 44 A1308238 tb26c11.x
c 28 267.8 15.3 271 40 AA918558 ol42f03.s
c 29 253.8 14.5 302 40 AA970909 oo82c05.s
c 30 249.8 14.3 441 21 T75407 yC90e12.r1
c 31 246.4 14.1 260 20 Z41879 HSC02A051 n
c 32 246.4 14.1 364 20 T08924 T08924 In
c 33 216 12.4 217 28 AA063085 zif6h05.s
c 34 207.8 11.9 395 69 AW125394 UI-N-BH2.
c 35 194.4 11.1 600 21 R12478 yf29c04.s1
c 36 186.2 10.7 197 28 AA063135 zif6h05.r
c 37 183 10.5 388 21 R09874 yf29c04.r1
c 38 180 10.3 189 38 AA758721 ah75f02.s
c 39 177 10.1 184 36 AA65039 nu70h05.s
c 40 173.4 9.9 410 21 R12815 yf57a12.r1
c 41 170.4 9.8 246 21 T89769 yd99d01.r1
c 42 169.2 9.7 182 23 H51212 yd33f05.s1
c 43 141.8 8.1 524 38 AA760180 vv74b01.r
c 44 140.6 8.0 454 33 AA455498 zx76h05.r
c 45 139.8 8.0 456 23 H60273 yr41e01.r1
```

## ALIGNMENTS

```

RESULT 1
AF150387 1114 bp mRNA EST 22-JUN-1999
LOCUS AF150387 Human mRNA from cd34+ stem cells Homo sapiens CDNA clone
DEFINITION CBMAPH03, mRNA sequence.
ACCESSION AF150387.1 GI:5133823
VERSION Est.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1114)
AUTHORS Ye,M., Zhang,Q.H., Zhou,J., Shen,Y., Wu,X.Y., Guan,Z.Q., Wang,L.,
Fan,H.Y., Mao,Y.F., Dai,M., Huang,Q.H., Chen,S.J. and Chen,Z.
TITLE Human mRNA from cd34+ stem cells
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3137104.
Contact: Ye M
Shanghai Institute of Hematology
Shanghai Second Medical University, Rui-Jin Hospital
197 Rui-Jin Road II, Shanghai, 200025, P. R. China
Email: zchen@stn.sh.cn.
FEATURES
Location/Qualifiers
source 1..1114
```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CBMAPH03"
/clone_lib="Human mRNA from cd34+ stem cells"
/tissue_type="umbilical cord blood"
/cell_type="cd34+ stem cell"
BASE COUNT 275 a 300 c 286 g 253 t
ORIGIN
Query Match 55.7%; Score 973.8; DB 51; Length 1114;
Best Local Similarity 97.0%; Pred. No. 5e-277;
Matches 1089; Conservative 0; Mismatches 22; Indels 12; Gaps 9;
QY 421 ggaagtatgacacactcgctgtacacagcagctctctgtgtatcacacccatccgcgaggag 480
DB 1 GGAGGATCGACCACTCGGTGTACAGAGCAGTCTCTGTATACACCCATCCGGCGGGAG 60
QY 481 tggattatgagtgatcatctgtgcgggtggagatcaatggacagagatctgaaatggac 540
DB 61 TGGTATTATGAGGTATCATCTTGTGCGGTGGAGATCAATGGACAGGATCTGAAATGGAC 120
QY 541 tgcaggagtacaactatgacaaagacatgtggacagctgggacagctgggacacacacctcttg 600
DB 121 TGCAGGAGTACAACTATGACAAGACATGTGGACAGTGGGACAGTGGCAAGCAGCAGCCCT 180
QY 601 cccaaagaagtgtttgaagctgcagtcacatccatcaagcagcctctccacagggagag 660
DB 181 CCCAAGAAAGTGTTCAGCTGAGTCAATCCATCAAGCAGCCTCTCCACGGAGAAG 240
QY 661 ttccctgcaggtttctggttaggagagcagctggtgtgtctggaagcagcagcaccct 720
DB 241 TTCCCTGATGTTTCTGGGTAGGAGAGCAGTGGTGTGTGGCAAGCAGCAGCAGCCCT 300
QY 721 tgaacattttccagctatctcaactcaatgaatgggtgaggttaccacacagctctc 780
DB 301 TGAACATTTTCCAGTCACTCTCACTTCTACCTAATGGGTAGGTATACCAACAGTCTTC 360
QY 781 cgatcacattcttcgcagcagaataccttgcgcagtggaagatgtggcagctgccca 840
DB 361 CGCATCACCATCTTCCGAGCAATACTTGGGCCA-TGGAAGATGTGGCAGCTGCCAA 419
QY 841 gacgactgtacaagtgtgcatctcacagctatcacagggcactgttatggagctgtt 900
DB 420 GACGACTGTATAC-AGTTTGCATCTCACAGTCACTCCACGGGCACTGTATGGAGCTGT 478
QY 901 atcatgaggggtctta-cgttctttgatcggtccggaacaaacgaattggttctgt 959
DB 479 ATCATGGAGGGCTTCTACCGTTCCTTTGATCGGGCCCCGAAACGAATTGGCTTGTGT 538
QY 960 cagcgttgcattgtgacagatgatttcaggagcggcagcgggtggaaggccntttgtca 1019
DB 539 CAGCGTGTGCATGTGCAGCATGAGTTCAGGACGGCAGCGGTGGAAGGCC--TTTGTC 596
QY 1020 ccttgacatggaagactgtggtacaacattccacagacagatgagtaaacctcatga 1079
DB 597 CCTTGCACATGGAAGACTGTGGCTCAACATTCACAGACAGATGATCAACCTCATGA 656
QY 1080 ccatagcattatgctggtgctgcatctgcgcctcttcttctgctgacctgctcatgg 1139
DB 657 CCATAGCCATATGTCATGGCTGCCATCTGCG-CTCTTTCATGCTGCCACTCTGCTCATGG 715
QY 1140 tgtgtcag-tggcgtgctcccgctgctgcgcgcagcagcatgatgacttggatgac 1198
DB 716 TGTGTGAGATGGGCTCCCTCCGCTGCTGCGCAGCAGCATGATGACTACTGTGTGATGAC 775
QY 1199 atctccctgctgagtgaggagggcccatggcagagataggattcccttgaccacac 1258
DB 776 ATCTCCCTGCTGAAGTGAGGAGGCCCATGGGAGAGATAGATATCCCTTGGACACAC 835
QY 1259 ctccgtggttcactttggtcacagttaggagacacagatggcacctgtggccagagacc 1318
DB 836 CTAGTGGTCACTTTGGCAC--AGTAGGAGACACAGATGGCACCCTGTGGCCAGAGACC 892
```



Matches 506; Conservative 0; Mismatches 3; Indels 1; Gaps 1

QY 1230 cagaagataggattccccctggaccacacacctccgtggttcacttttggtcacaagttagag 1289  
DB 525 CAGAAGATAGAGATTCCCTGGACCACACCTCCGTGTTCACTTGTTCACAAGTAGGAG 466  
QY 1290 acacagatggcaccttggcccagagacacctcaggaacctccccaccacacaaaatgcctct 1349  
DB 465 ACACAGATGGCACCTGTGGCCAGAGACACTCAGAGACCTGCCACACCCACCAAATGCCTCT 406  
QY 1350 gcttgtatgagaagaaaaagcgctggcaagtggtgtccaggagactgacctgtaggaga 1409  
DB 405 GCCTTGATGGAGAAAGGAAGCGTGGCAAGTGGGTGCCAGGAGCTGTACTGTAGGAAA 346  
QY 1410 cagaaagagaagaaaaaacctctgctggcggaatactcttggcacctcaaatatta 1469  
DB 345 CAGAAAAGAGAAAGAACGACTCTGCTGGCGGAATACTCTTGGTCACCTCAAAATTA 286  
QY 1470 agtcggaaattctgctgttgtaaacttcagccctgaacctttgt-cacacttcccttaa 1528  
DB 285 AGTCGGGAAATTCGTGCTTGAACCTCAGCCCTGAACCTTGTCCACCAATTCCTTTAA 226  
QY 1529 attctcaacccaagattctctttcttttagtttcagaagtactggcatcacacgag 1588  
DB 225 ATTCTCAACCCAAGATTCTCTTTCTTAGTTTTCAGAAAGTACTGGCATCACAGGAG 166  
QY 1589 gttaacctggcgtgtgccctgtgtagccctggcagagaagaccaaagctgtttccct 1648  
DB 165 GTTACCTTGGCGTGTGTCCTGTGGTACCTGGCAGAGAGACCAAGCTTGTTCCT 106  
QY 1649 gctggccaaagtctagtgaggagagatgcacagttgttatgttttagagacagggact 1708  
DB 105 GCTGGCCAAAGTCAGTAGGAGAGGATGCACAGTTGTGTTATTTGCTTTAGACAGGGGACT 46  
QY 1709 gtataaacaagcctaactggtgtcaaaa 1738  
DB 45 GTATAAACAAGCCTAACATTGGTGCAAGA 16

RESULT 3  
AA701598/c

LOCUS AA701598 464 bp mRNA EST 19-DEC-1997  
DEFINITION z134n07.s1 Soares\_fetal\_liver\_spleen\_INFLS\_S1 Homo sapiens CDNA clone IMAGE:432733 3', mRNA sequence.  
ACCESSION AA701598  
VERSION AA701598.1 GI:2704763  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 464)  
AUTHORS Hillier,L., Allten,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wyllie,T., Waterston,R. and Wilson,R.  
TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1392818.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
..This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40n13 fwd. ET from Amersham.  
Location/Qualifiers  
FEATURES  
SOURCE 1..464  
/organism="Homo sapiens"  
/db\_xref="GDB:1332502"  
/db\_xref="taxon:9606"

```
/clone="IMAGE:432733"
/clone_lib="Soares_fetal_liver_spleen_INFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5',
RACGTGAGAAATTAATTAAGATCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      115 a 110 c 114 g 125 t
ORIGIN
Query Match      24.9%; Score 434.8; DB 37; Length 464;
Best Local Similarity 99.3%; Pred. No. 1e-117;
Matches 447; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1290 acacagatggcactgtgcccagagcacctcaggaccctccccaccacccaatgcctct 1349
|||||
Db 464 ACACAGATGGCACCCTGTGGCCAGAGCACCTCAGGACCCTCCGCCACCACCAATGCCCTCT 405
QY 1350 gcttgatggagaagaaaagctgccaagtggtggtccaggacctgtacctgtaggaga 1409
|||||
Db 404 GCCTTGATGGAGAAGAAAGGCTGCCAAGGTGGGTTCAGGACACTGTACCTGTAGGAAA 345
QY 1410 cagaagaagaagaagcactctgctggcggaatactctgtccaccctcaaatTTA 1469
|||||
Db 344 CAGAAAAGAGAGAAGACACTCTGTGGCGGAATACTCTGTGTACCTCAAAATTTA 285
QY 1470 agtcgggaattctgtgtgaaacttcagocctgaaccttgtt-caccattcccttaa 1528
|||||
Db 284 AGTCGGGAAATTCCTGCTGTTGAAACTTCAGCCCTGAACCTTTGTCCACCATTCTTTAA 225
QY 1529 attctccaccacaaagtattcttcttagtttcagaagctactggcattcacacgcag 1588
|||||
Db 224 ATTCTCCACCCAAAGATATCTCTTTCTTAGTTTCAGAAGTACTTGGCATCACAGCAG 165
QY 1589 gttaccttggtgtgtccctgtgtgacctgtggcagagaagaccaaagcttgttccct 1648
|||||
Db 164 GTTACCTTGGCGTGTCCCTGTGGTACCTTGCAGAGAAGACCAAGCTTGTTCCT 105
QY 1649 gctggccaaagtccagtaggagagatgcacagtttctatttcttagacacagggact 1708
Db 104 GCTGGCCAAAGTCAGTAGGAGAGATGCACAGTTTCTATTGCTTTTAGAGACAGGACT 45
QY 1709 gtataaacaagcctaacattggtgcaaaaa 1738
Db 44 GTATAACAAGCCTAACATTTGGTGCAAGA 15
RESULT 4
AI127789/c 464 bp mRNA EST 27-OCT-1998
LOCUS q33c02.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
DEFINITION IMAGE:1711298 3', mRNA sequence.
ACCESSION AI127789
VERSION AI127789.1 GI:3596303
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 464)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
```

```
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2152649.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 804 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 445.
FEATURES
Location/Qualifiers
source
1..464
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1711298"
/clone_lib="Soares_pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AATCGGAAGAAATTCGCGCGCGCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
BASE COUNT      115 a 110 c 114 g 125 t
ORIGIN
Query Match      24.9%; Score 434.8; DB 42; Length 464;
Best Local Similarity 99.3%; Pred. No. 1e-117;
Matches 447; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1290 acacagatggcactgtgcccagagcacctcaggaccctccccaccacccaatgcctct 1349
|||||
Db 464 ACACAGATGGCACCCTGTGGCCAGAGCACCTCAGGACCCTCCGCCACCACCAATGCCCTCT 405
QY 1350 gcttgatggagaagaaaagctgccaagtggtggtccaggacctgtacctgtaggaga 1409
|||||
Db 404 GCCTTGATGGAGAAGAAAGGCTGCCAAGGTGGGTTCAGGACACTGTACCTGTAGGAAA 345
QY 1410 cagaagaagaagaagcactctgctggcggaatactctgtgacctcaaatTTA 1469
|||||
Db 344 CAGAAAAGAGAGAAGACACTCTGTGGCGGAATACTCTGTGTACCTCAAAATTTA 285
QY 1470 agtcgggaattctgtgtgaaacttcagocctgaaccttgtt-caccattcccttaa 1528
|||||
Db 284 AGTCGGGAAATTCCTGCTGTTGAAACTTCAGCCCTGAACCTTTGTCCACCATTCTTTAA 225
QY 1529 attctccaccacaaagtattcttcttagtttcagaagctactggcattcacacgcag 1588
|||||
Db 224 ATTCTCCACCCAAAGATATCTCTTTCTTAGTTTCAGAAGTACTTGGCATCACAGCAG 165
QY 1589 gttaccttggtgtgtccctgtgtgacctgtggcagagaagaccaaagcttgttccct 1648
|||||
Db 164 GTTACCTTGGCGTGTCCCTGTGGTACCTTGCAGAGAAGACCAAGCTTGTTCCT 105
QY 1649 gctggccaaagtccagtaggagagatgcacagtttctatttcttagacacagggact 1708
Db 104 GCTGGCCAAAGTCAGTAGGAGAGATGCACAGTTTCTATTGCTTTTAGAGACAGGACT 45
QY 1709 gtataaacaagcctaacattggtgcaaaaa 1738
Db 44 GTATAACAAGCCTAACATTTGGTGCAAGA 15
RESULT 5
AI127822/c 443 bp mRNA EST 27-OCT-1998
LOCUS q336f02.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
DEFINITION
```

[illegible]



subtracted NCI\_CGAP libraries: NCI\_CGAP\_Co4, NCI\_CGAP\_Pr28, NCI\_CGAP\_Co10, NCI\_CGAP\_Co16, NCI\_CGAP\_Kid3, NCI\_CGAP\_Kid12, NCI\_CGAP\_Kid3, NCI\_CGAP\_Kid11, NCI\_CGAP\_Lym2, NCI\_CGAP\_Br2, NCI\_CGAP\_Co8, NCI\_CGAP\_CLL1, NCI\_CGAP\_Le12, NCI\_CGAP\_Br23, NCI\_CGAP\_Lu5, NCI\_CGAP\_Lu24, NCI\_CGAP\_Lu19, NCI\_CGAP\_G04, NCI\_CGAP\_G05, NCI\_CGAP\_Br25. These 21 libraries were pooled and a single-stranded DNA preparation of the resulting mixture was used as a tracer in a subtractive hybridization with a driver whose composition is detailed below: NCI\_CGAP\_Kid3 pool 1 LLAM 3334-3337, 3682-3683, 3798-3803 (IMAGE Clonides) 1322376-1323911, 1456008-1456775, 1500552-1502855) NCI\_CGAP\_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonides 1323912-1325831, 1471368-1472903, 1492104-1493255) NCI\_CGAP\_Lu5 pool 1 LLAM 3575-3582, 3851-3854 (IMAGE Clonides 1414920-1417991, 1520904-1522439) NCI\_CGAP\_G04 pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE Clonides 1257096-1258631, 1469064-1470983, 1475592-1476743) NCI\_CGAP\_Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonides 985608-986759, 1101192-1101959, 1217928-1220615) NCI\_CGAP\_Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE Clonides 1057416-1061255, 1144584-1145351) The resulting subtracted library contained 530,000 recombinants. Subtraction was performed as previously described. [Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806. TAG\_LIB=NCI\_CGAP\_Kid3 TAG\_TISSUE=kidney TAG\_SEQ=AAATGC

BASE COUNT 111 a 98 c 105 g 133 t  
ORIGIN  
Query Match 22.6%; Score 394.4; DB 64; Length 447;  
Best Local Similarity 98.3%; Pred. No. 8.9e-106;  
Matches 409; Conservative 0; Mismatches 6; Indels 1; Gaps 1;  
QY 1324 accctcccccaccacaaatgctctgcttgatgagagaaagggctggcaagtggtg 1383  
|||||  
Db 447 ACCCTCCCAACCACCAATGCTCTGCTTGTGAGAGAGAAAGGCTGGCAGAGTGG 388  
QY 1384 gttccaggagactgtacctgttagagacagaaagagaaagacactctgctggcgg 1443  
|||||  
Db 387 GTTCCAGGAGCTGTACCTGTAGGAAACAGAAAGGAGAAAGAGCACTCTCTGGCGG 328  
QY 1444 gaatactcttggtaccccaatttaagtcgggaaattctgtctgtgtaacttcagccc 1503  
|||||  
Db 327 GAATACTCTGTGTACCTCAATTAAGTCGGGAAATTCGTGTGTAACCTTCAGCCC 268  
QY 1504 tgaacctttgt-cacattctcttaattctcaaccacaaagtattctctttcttagt 1562  
|||||  
Db 267 TGAACCTTTGTCCACCATTCCTTAATTCCTCAACCCCAAGATATCTCTTTCTTACT 208  
QY 1563 ttcaagaagtactggcatcacacgcagggttaccttggtgtgtgtccctgtggtaccctggc 1622  
|||||  
Db 207 TTCAGAAAGTACTGGCATCACGCAGGTTACCTTGGCGGTGTCCCTGTGGTACCCAGCG 148  
QY 1623 agagaagagaccacagctgtttccctgtgtggccaaagtacgtaggagagatgcacagtt 1682  
|||||  
Db 147 AGAGAAGAGACCAAGCTTTTCCCTGTGTCGCCCAAGTCAAGTAGGAGAGGATGCACAGTT 88  
QY 1683 tgcattttcttagagacagggactgtataacagccctaacattggtggtcaaaaa 1738  
|||||  
Db 87 TCGTATTTCCTTTAGAGACAGGGACTGTATAACACAGCCCTACATTTGGTGCAAGA 32  
RESULT 9  
AA677847/c 427 bp mRNA EST 19-DEC-1997  
LOCUS z113c06.s1 Soares\_fetal\_liver\_spleen\_1NFLS\_S1 Homo sapiens cDNA  
DEFINITION

Clone IMAGE:430666 3', mRNA sequence.  
AA677847  
VERSION AA677847.1 GI:2658369  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 427)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On Nov 6, 1997 this sequence version replaced gi:1565841.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40ml3 fwd. ET from Amersham.  
FEATURES  
Location/Qualifiers  
1..427  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:430666"  
/clone\_lib="Soares\_fetal\_liver\_spleen\_1NFLS\_S1"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen 1NFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGATTAATTAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."  
BASE COUNT 111 a 99 c 102 g 115 t  
ORIGIN

Query Match 22.1%; Score 386.8; DB 37; Length 427;  
Best Local Similarity 99.0%; Pred. No. 1.5e-103;  
Matches 410; Conservative 0; Mismatches 2; Indels 2; Gaps 2;  
QY 1326 cctcccccaccacaaatgctctgcttgatgagagaaagggctggcaagtggtg 1385  
|||||  
Db 427 CCTCCCCACCACCAATGCTCTGCTGTGAGAGAGAAAGGCTGGCAAGTGGGT 368  
QY 1386 tccaggagactgtacctgttaggagacagaaagagaaagacactctgtctggcgga 1445  
|||||  
Db 367 TCCAGGAGCTGTACCTGTAGGAAACAGAAAGAGAAAGACACTCTGCTGGCGGA 308  
QY 1446 atactctgggtacactcaaatgaagtcgggaaattctgtctgtgaaacttcagccctg 1505  
|||||  
Db 307 ATACTCTTT-GTCACTCAAAATTAAGTCGGGAAATTCGTCTTGAACACTTCAGCCCTG 249  
QY 1506 accctttgt-caccattcttaattctcaaccacagattctcttcttagttt 1564  
|||||  
Db 248 AACCTTTGTCCACCATCTTTAAATTCCTCAACCCCAAGTATCTCTTTCTTAGTTT 189  
QY 1565 cagaagtactggcatcacacgcagggttaccttggcgtgtgtccctgtgtgtaccctggcag 1624  
|||||  
Db 188 CAGAAGTACTGGCATCACACGCGAGGTACCTTGGCGGTGTGCCCTGTGGTACCTGGCAG 129

QY 1625 aaagagacacacgtgtttcctgctgcccgaagtcagtagcagagagatgcacagttag 1684  
 Db 128 AGAAGAGACCAAGCTTGTTCCTGCTGGCCAAAGTCAGTAGGAGAGGATGCACAGTTTG 69

QY 1685 ctattgtcttagacagagagctgtataaacagcctaactgtggtgcaaaa 1738  
 Db 68 CTATTGCTTTAGACAGAGGAGCTGTATTAACAAGCTTACATGTTGCAAGA 15

RESULT 10  
 R55298/c 448 bp mRNA EST 22-MAY-1995  
 LOCUS  
 DEFINITION: v177e01.sl-Soares breast 2NDBst Homo sapiens cDNA clone

ACCESSION R55298  
 VERSION R55298.1 GI:824593  
 KEYWORDS EST.  
 SOURCE human

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 448)  
 AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
 Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
 Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
 Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
 Wilson,R.

TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

Insert Size: 1306  
 High quality sequence stops: 339 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 1306 Std Error: 0.00  
 Seq primer: Promega -2lm3  
 High quality sequence stop: 339.

FEATURES  
 source  
 1. 448  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="GDB:556877"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Soares breast 2NDBst"  
 /sex="Female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: breast; Vector: pT7T3D (Pharmacia) with a  
 modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTACCAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTT 3']  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of a modified pT7T3 vector (Pharmacia).  
 Library went through one round of normalization to a Cot =  
 230. Library constructed by Bento Soares and M. Fatima  
 Bonaldo."

BASE COUNT 108 a 100 c 111 g 122 t 7 others  
 ORIGIN

Query Match 21.7%; Score 379.2; DB 22; Length 448;  
 Best Local Similarity 94.7%; Pred. No. 2.8e-101;  
 Matches 410; Conservative 0; Mismatches 20; Indels 3; Gaps 2;

QY 1309 ccagagacacctcagacc--tccccaccacacaaatgcctctgctgtaggagaagga 1366  
 Db 447 CCAAACNCTCAGGACCCCTCCCAACCNACCAAGCCTCTGCCTGTATGGAGAGNA 388

QY 1367 aaagctgcgaagtggttcagggactgtacctgttaggagacagaaagagaagaag 1426  
 Db 387 AAAGGCTGGCAAGGTGGGTGCCGAGGACTGTACCTGTAGGAACAGAAAGAGAGAAG 328  
 QY 1427 aagcactctgtgcggggaataactcttggctcacctcaaatatttaagtcgggaaattcgtct 1486  
 Db 327 AAGCACTCTGCTGGCGGAATACTCTTGGTCACCTCAAAATTTAAGTCGGGAANTCTGCT 268  
 QY 1487 gtttgaacctcagccctgaacctttgt-caaccattcctttaaatttcccaacccaaagt 1545  
 Db 267 GCTTGAACCTTCAGCCCTGAACCTTTGTCCACCATTCCTTTAAATTCCTCAACCCAAAGT 208  
 QY 1546 attctcttttcttagtttcagaagtactggcatcacacgcaggttaccttggcgtgtgt 1605  
 Db 207 ATCTCTCTTTCTTAGTTTCAGAAGTACTGGCATCACACGCAAGTTTACCTTGGCGTGT 148  
 QY 1606 cccgtgttacccctggcagagagacaaagctgtttccctctgtctgcccgaagtcagta 1665  
 Db 147 CCTGTGGTACCCAGCAGAGAGAGACCAAGCTTGTTCCTGCTGGCCAAAGTCAGTA 88  
 QY 1666 ggagagatgcacagtttgcattgttcttagacagagagactgtataaacagcctaac 1725  
 Db 87 GGAGAGGNTGCACAGTTTCTATTGCTTTAGACAGAGGAGCTGTATTAACAAGCCTAAC 28  
 QY 1726 attggtgcaaaa 1738  
 Db 27 ATTGGTGCAANA 15

RESULT 11  
 AI248727/c

LOCUS  
 DEFINITION

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE

ORGANISM  
 Homo sapiens

REFERENCE  
 AUTHORS  
 TITLE

JOURNAL  
 COMMENT

AI248727  
 qh72d06.x1 Soares fetal\_liver\_spleen\_1NFLS\_S1 Homo sapiens cDNA  
 clone IMAGE:1850219 3', mRNA sequence.

AI248727  
 EST.  
 human.

AI248727.1 GI:3844124  
 human.

Human sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 449 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 336.

Location/Qualifiers  
 1. 411  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:1850219"  
 /clone\_lib="Soares\_fetal\_liver\_spleen\_1NFLS\_S1"  
 /sex="male"

FEATURES  
 source

1. 411  
 /dev\_stage="20 week-post conception fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
 with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;  
 This is a subtracted version of the original Soares fetal  
 liver spleen 1NFLS library. 1st strand cDNA was primed  
 with a Pac I - oligo(dT) primer [5'  
 AACTGGAAGATTAATAAGATCTTTTTTTTTTTTTTTTTTTT 3']  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Pac I and cloned into the Pac I  
 and Eco RI sites of the modified pT7T3 vector. Library



	BASE COUNT	105 a	97 c	89 g	107 t	
ORIGIN						
Query Match	20.9%;	Score 365.4;	DB 41;	Length 398;		
Best Local Similarity	99.5%;	Pred. No. 3.2e-97;				
Matches 377;	Conservative	0;	Mismatches 1;	Indels 1;	Gaps 1;	
Qy	1356	atggaagaagaaaggctggcaaggtgggtttccagggaactaccttagagacagaaaa	1415			
Db	398	ATGGAAGAAGAAAAGCGTGGCAAGGTGGGTTCACAGGGACTACTCTGAGAAACAGAAA	339			
Qy	1416	agagaagaagaagcactctgtgcggggaatactcttggtcacctcaaattaagtctgg	1475			
Db	338	AGAGAAGAAGAAGACACTCTGCTGGCGGGANTACTCTTGGTCACCTCAAAATTAAGTCGG	279			
Qy	1476	gaattctgctgcttgaaacttcagccctgaacctttgttgcacattcccttaaatcttc	1534			
Db	278	GAAATTCCTGCTTGCTTGAAACTTCAGCCCTGAACCTTTGTCACCACTTCCTTTAAATTC	219			
Qy	1535	caaccacaagtatctctctttcttagtttcagaagttactggcatcacacagagttacc	1594			
Db	218	CAACCCAAGATATCTCTCTTTTCTTAGTTTCAGAAGTACTGGCATCACACGAGGTTACC	159			
Qy	1595	tggcggtgtgctcctgtggtacccttggcagagaagagaccaagctgtgttccctcgtctgpc	1654			
Db	158	TTGGCGCTGTGCTCCTGTGGTACCTCTGGCAGAGAAGACCAAGCTGTGTTCCCTCTGCTGGC	99			



Search completed: April 30, 2000, 13:28:01  
Job time: 10661 sec

```

FEATURES
source
High quality sequence stop: 414.
Location/Qualifiers
1. .420
/organism="Homo sapiens"
/db_xref="CDB:421950"
/db_xref="taxon:9606"
/clone IMAGE:49409
/clone_lib="Soares infant brain lNIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lfamid BA; Site_1: Not I; Site_2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'; AACTGGAGAAATTCGGCGCCGAGGAATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lfamid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      98 a    116 c    107 g     95 t          4 others
ORIGIN
Query Match           19.7%; Score 343.8; DB 22; Length 420;
Best Local Similarity 94.1%; Pred. No. 8.le-91;
Matches 398; Conservative 0; Mismatches 21; Indels 4; Gaps 4;
Qy 1105 ctgcgcocttcatgtcgccactctgcoctcatgtgtgtcagtgcgctgcctccgctg 1164
|||||

```

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 10:58:20 ; Search time 4425.31 Seconds  
(without alignments)  
-1198.687 Million cell updates/sec

**Title:** US-09-215-435-66

Perfect score: 1747

Sequence: 1 gcctaccaatggttcctt.....tggtgcaaaaaaaaaa 1747

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 821193 seqs, -1518192014 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:★

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_on:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_p1:\*  
8: gb\_p12:\*  
9: gb\_p1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_v1:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_on:\*  
22: em\_or:\*  
23: em\_pat:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_to:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_v1:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: gb\_in3:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*

```
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: qb_pl3:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1267	72.5	5051	10	AB032975	AB032975 Homo sapi
2	1232	70.5	2526	10	AF190725	AF190725 Homo sapi
3	1232	70.5	3252	40	AF200193	AF200193 Homo sapi
C 4	452	25.9	176874	32	AF000761	AF000761 Homo sapi
C 5	350	20.0	108772	32	AP000685	AP000685 Homo sapi
C 6	180	10.3	600	13	G24698	G24698 human STS W
7	53	3.0	2158	12	AF190727	AF190727 Rattus no
8	53	3.0	4001	12	AF190726	AF190726 Mus muscu
C 9	23	1.3	122568	40	AF196972	AF196972 Homo sapi
10	22	1.3	50010	43	AC034457	AC034457 Drosophill
11	22	1.3	121043	33	AC007710	AC007710 Drosophill
12	22	1.3	154469	11	HSJ729N16	AL096707 Human DNA
13	21	1.2	972	12	AF051150	AF051150 Mus muscu
14	21	1.2	23547	45	AC017979	AC017979 Drosophill
15	21	1.2	31985	43	AC013831	AC013831 Drosophill
16	21	1.2	59003	44	AC013254	AC013254 Drosophill
17	21	1.2	60956	43	AC013689	AC013689 Homo sapi
C 18	21	1.2	104033	11	HS751H9	AL034377 Human DNA
19	21	1.2	108881	8	ATAC002391	AC002391 Arabidops
C 20	21	1.2	110000	45	AC012152_1	Continuation (2 of
C 21	21	1.2	110054	8	F7G19	AC00106 Sequence
22	21	1.2	123189	11	AC002556	AC002556 Human Chr
C 23	21	1.2	133981	41	AC009742	AC009742 Drosophill
C 24	21	1.2	166331	44	AC013588	AC013588 Homo sapi
25	21	1.2	203488	42	AC009488	AC009488 Homo sapi
C 26	21	1.2	227968	40	AF053356	AF053356 Homo sapi
27	20	1.1	220	10	HUMPEP5	J00283 Human pepsi
C 28	20	1.1	250	13	G15147	G15147 human STS S
29	20	1.1	293	9	HUMPEPA05	M26029 Human pepsi
C 30	20	1.1	375	35	AF015472	AF015472 plasmodiu
C 31	20	1.1	774	12	AB006880	AB006880 Rattus no
32	20	1.1	822	12	AF144701	AF144701 Rattus no
33	20	1.1	1338	9	MACPGA	M20788 Rhesus monk
34	20	1.1	1353	10	MFPEPA23	X59755 M.fuscata m
C 35	20	1.1	1620	12	RNTG05	X05145 Rat TO gene
36	20	1.1	2470	10	MFPEPSA1	X39752 M.fuscata g
37	20	1.1	3051	4	XLU029178	U29178 xenopus lae
38	20	1.1	5880	43	AC014020	AC014020 Drosophill
39	20	1.1	8233	12	MMBA341A43	U41396 mus musculu
C 40	20	1.1	22468	35	AF068693	AF068693 Trypanoso
41	20	1.1	45851	43	AC012991	AC012991 Drosophill
42	20	1.1	50276	41	AC009746	AC009746 Drosophill
43	20	1.1	6167	11	HSBAB9B2	AL080245 Human DNA
44	20	1.1	71645	41	AC011689	AC011689 Homo sapi
C 45	20	1.1	76713	8	ATAC006437	AC006437 Arabidops

## ALIGNMENTS

RESULT	1
AB032975	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	

PRI 11-NOV-1999  
partial cds.

AB032973  
AB032975.1 GT:6330044

KEYWORDS Homo sapiens adult male brain cDNA to mRNA, clone\_lib:pBluescriptII  
 SOURCE SK plus clone:hg01289.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Hirose, M., Nagase, T., Ishikawa, K., Kikuno, R., Nomura, N. and  
 Chara, O.  
 TITLE Characterization of cDNA clones selected by the Genemark analysis  
 from size-fractionated cDNA libraries from human brain  
 JOURNAL DNA Res. 6 (5), 329-336 (1999)  
 MEDLINE 20039618  
 REFERENCE 2 (bases 1 to 5051)  
 AUTHORS Ohara, O., Nagase, T. and Kikuno, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-OCT-1999) to the DDBJ/EMBL/GenBank databases. Osamu  
 Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology,  
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: cdmain@kazusa.or.jp, URL: http://www.kazusa.or.jp/huge/,  
 Tel.: +81-438-52-3913, Fax: +81-438-52-3914)  
 FEATURES  
 Location/Qualifiers  
 1..5051  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="hg01289"  
 /clone\_lib="pBluescriptII SK plus"  
 /dev\_stage="adult"  
 /sex="male"  
 /tissue\_type="brain"  
 1..1191  
 /gene="KIAA1149"  
 1..1191  
 /gene="KIAA1149"  
 /note="Start codon is not identified."  
 /codon\_start=1  
 /product="KIAA1149 protein"  
 /protein\_id="BAA86463.1"  
 /db\_xref="gi:6330045"  
 /translation="YLQAHFTFICSGWSYVDLRKGVVYPTGKVEGELGDLVSP  
 HGNVTVANTAAITSEDKFFICNSNWEGLIGLAYEARPPDSLEFFDLSLVKQTHV  
 NPLFSLOICAGDFPLNSELVASGVSGMITIGIDHSLYTGLWYTPIRREMYEYVIL  
 RVEINLODKMDCKEYNDKSIYDGTNLRPKKVFEEAAVKSIKAASTEKFPDGFV  
 LGEQLVQACQATTPWNIPIVLSYLMGEVNTNOSFRITLPQOYLRPVEDVATSDQDQ  
 KFAISQSSITGVMGAVINVEGVVDFRARKRIGFAVSACHVDFEFAAEGVGFVIL  
 MEDCGNIPQIDESTLITATVMAAICAFMLPLCLMWQVRCRLRQLRQDHDADFADI  
 SLK"  
 BASE COUNT 1385 a 1207 c 1137 g 1322 t  
 ORIGIN  
 Query Match 72.5%; Score 1267; DB 10; Length 5051;  
 Best Local Similarity 99.6%; Pred. NO. 0;  
 Matches 1707; Conservative 0; Mismatches 4; Indels 2; Gaps 2;  
 Yy 25 tatgtgcaagccactttacactctgtcttgggtgtccagcacataccggacactccgg 84  
 Db 1 TATCTGAGCCCATTTTACACTCTGTTCTGGTGTGTCAGCACATACCGGGACCTCGG 60  
 Yy 85 aagsggtgtatgtgcctcacaccagggcaagtggaaaggagctgggaccgacctg 144  
 Db 61 AAGGGTGTGTATGTGTCCTACACCGGGCAAGTGGAGGGAGCTGGGCCACCGACTG 120  
 Yy 145 gtaagcatccccatggccccaaactgactgtgcgtgccaaactgtgcctacactgaa 204  
 Db 121 GTAGCATCCCCATGCCCCCAAGCTCAGTGTGCGTGCCCAACATGTCGCATCACTGAA 180  
 Yy 205 tcagacaagttcttcataacaggtctcaactgggaaggtactctgggctgggctatgct 264  
 Db 181 TCAGACAAGTTCCTCATCAACGGGTCCCACTGGGAAGGACATCTCGGGCTGGGCTATGCT 240  
 Yy 265 gagattgcaggcctgcagactccccggagcctcttcttactctctgtgtaagcagacc 324





Db	972	TTGGAACATTTTCCAGTCTACTCTACCTAATGGTGGAGGTTACCAACACCTCCTT	1031
QY	780	ccgcatcaccatctctccgacataactctgcygcagtggaagatggtggccacatcccca	839
Db	1032	CCGCATCACCATCTTCCGACAACTACTGGCGCAGTGGGAAGATGGCCACATCCCA	1091
QY	840	agacactgttacaaagtgtgcatctccacagtcacccacgagcactgtatggagctgt	899
Db	1092	AGACACCTGTACAAAGTTGCCATCTCACAGTCATCCAGGCACTGTATGGAGCTGT	1151
QY	900	tatcatgagggcttctacgttctgttcttgatcgggccgaaacgaattggttctgt	959
Db	1152	TATCATGAGGGCTTCTAGCTGTCTTGTATCGGCGCCGAAACGAATGGCTTTCGT	1211
QY	960	cagccttgccatgtgcacatgagtcaagcgcagcgcggtggaagccentttgtca	1019
Db	1212	CAGCCCTTGCATGTGCAGATGAGTTCAGGACGCGGCGGGAAGGCCCTTTTGTCA	1270
QY	1020	ccttgacatgaagactgtgctacacattccacagacagatgagtcacccatcatga	1079
Db	1271	CCTTGACATGAAGACTGTGGCTACAACTTCCACAGACATGAGTCAACCCCTCATGA	1330
QY	1080	ccatgacctatgtcgtggtccatctgcgcctcttcatgtgccaactctgcctcatgg	1139
Db	1331	COATAGCCTATGTCTATGGCTGCCATCTGGCCCTCTTCTATGTCGCATCTGCTTCATGG	1390
QY	1140	tgtgtcagtggtcgtcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt	1199
Db	1391	TGTGTCACTGGCGTCCCTCCCTGCTGCGCCAGCAGCATGATGCTTGTGTATGACA	1450
QY	1200	tctcctctgaagtgaagagccatcggcagagatgaggtccctgagacacacc	1259
Db	1451	TCTCCCTGCTGAAGTGAGAGGCCATCGGCGAGATAGATGATTCCTCGGACACACCT	1510
QY	1260	tcgtggttccacttgggtcagtagtagacacagagagagagagagagagagagagag	1319
Db	1511	TCCGTGGTTCATTTGGTTCACAAAGTAGGAGACACAGATGGCACCTGTGGCAGACCT	1570
QY	1320	caggccctcccccacccacccatcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt	1379
Db	1571	CAGGACCTCCCGACCCACCAATTCCTGCTGCTGATGGAGAGAGAGAGAGAGAGAG	1630
QY	1380	gtgggtccaggtactgtacgttaggagacagagagagagagagagagagagagagag	1439
Db	1631	GTGGGTTCCAGGAGTGTACCTGTAGGAACAGAGAGAGAGAGAGAGAGAGAGAGAG	1690
QY	1440	gcgggaatactctgtgacaccccaatttaagtcgaggaattctcgtcgtcgtcgtcgtcgt	1499
Db	1691	GGGGAAATACCTTGTGTACCTCAAAATTAAGTCGGGAAATTCGTCTGCTGAACATCA	1750
QY	1500	gcctgaaccttgtt-caccatctcttaaatctccacaccccaagattctctctctct	1558
Db	1751	GCCCTGAACCTTGTCCACCATCTTAAATTCCTCAACCCCAAGAGTATCTCTTCTTCT	1810
QY	1559	tagttcagaagtactgcatcacacgaggttactcgtggtcgtcgtcgtcgtcgtcgtcgt	1618
Db	1811	TAGTTTCAGAACTGTGGATCACAGCAGGTACTTTCCTGCGTGTCTCCCTGTGTACCC	1870
QY	1619	tgcgagaagagacacagcttcttccctgctggtcccaagtcagtaggagagatgcac	1678
Db	1871	TGGCAGAGAAGACCAAGCTTTCCTGCTGCTGGCCAAAGTCAGTAGGAGAGATGCAC	1930
QY	1679	agtttgcatttctgttagagacagggactgtataaacaagcctaacattggtgcaaa	1736
Db	1931	AGTTTGTCTATTGCTTTAGACAGGGAGTGTATAAACAAGCTTAACATTTGTGTGAAA	1988
RESULT	4		
AP000761/c			
LOCUS			
DEFINITION	AP000761, 176874 bp DNA	HTG	27-NOV-1999
SEQUENCING	Homo sapiens chromosome 11 clone RP11-794I11 map 11q23, ***		
ACCESSION	SEQUENCING IN PROGRESS ***, in unordered pieces.		
	AP000761		

VERSION	AP000761.1	GI:6469067
KEYWORDS	HTG; HTGS-PHASE1.	
SOURCE	Homo sapiens DNA, clone:RP11-794I11.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 176874) Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.	
TITLE	Homo sapiens 176,874 genomic DNA of 11q23	
JOURNAL	Published Only in Database (1999) In press	
REFERENCE	2 (bases 1 to 176874) Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.	
AUTHORS	Direct Submission	
TITLE	Submitted (25-NOV-1999) to the DDBJ/EMBL/GenBank databases.	
JOURNAL	Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan	
COMMENT	(E-mail:hattori@gs.c.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923, Fax:81-42-778-9924) NOTE: This is a 'working draft' sequence. It currently consists of 41 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved	
	1	17060 contig of 17060 in length
	17561	26431 contig of 8871 in length
	26932	37174 contig of 10243 in length
	37675	50315 contig of 12641 in length
	50816	59520 contig of 8705 in length
	60021	66674 contig of 6654 in length
	67175	72749 contig of 5575 in length
	73250	79383 contig of 6134 in length
	79884	85508 contig of 5625 in length
	86009	91395 contig of 5387 in length
	91896	95164 contig of 3269 in length
	95665	99111 contig of 3447 in length
	99612	103934 contig of 4323 in length
	10435	108668 contig of 4234 in length
	109169	112967 contig of 3799 in length
	113468	116560 contig of 3093 in length
	120740	120239 contig of 2653 in length
	123893	127142 contig of 3250 in length
	127643	130250 contig of 2608 in length
	130751	132561 contig of 1811 in length
	133062	135186 contig of 2125 in length
	135687	137591 contig of 1905 in length
	138092	139516 contig of 1425 in length
	140017	141828 contig of 1812 in length
	142329	144664 contig of 2336 in length
	145165	147389 contig of 2225 in length
	147890	149432 contig of 1543 in length
	149933	151748 contig of 1816 in length
	152249	154238 contig of 1990 in length
	154739	156951 contig of 2213 in length
	157452	158677 contig of 1226 in length
	159178	161160 contig of 1983 in length
	161661	163453 contig of 1793 in length
	163954	165709 contig of 1756 in length
	166210	167786 contig of 1577 in length
	168287	169682 contig of 1396 in length
	170183	171628 contig of 1446 in length
	172129	173191 contig of 1053 in length
	173692	175227 contig of 1536 in length
	175728	176874 contig of 1147 in length.
		* NOTE: This is a 'working draft' sequence.
		* This record will be updated with the finished sequence
		* as soon as it is available and the accession number will
		* be preserved.

FEATURES		Location/Qualifiers	
source		1. .176874	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/chromosome="11"	
		/clone="RP11-794111"	
		/map="11q23"	
BASE COUNT		40376 a 36649 c 35725 g 40293 t 23831 others	
ORIGIN			
Query Match		25.9%; Score 452; DB 32; Length 176874;	
Best Local Similarity		99.4%; Pred. No. 6e-253;	
Matches 722; Conservative		0; Mismatches 3; Indels 1; Gaps 1;	
QY	1012	ttttgtcacttgacatggaagactgtggtacacattccacacagatgagtcac 1071	
Db	83997	TTTTGTCACTTGGACATGGAAGACTGTGGCTACAACTTCCACAGATGAGTCAAC 83998	
QY	1072	cctcatgacatagcctatgtcatggtgctccatctgccccttctcatgctgaccttg 1131	
Db	83937	CCTCATGACCATAGCCTATGTCATGCTGCCATCTGGCCCTCTTCATGCTGCACCTCG 83938	
QY	1132	cctcatggtgtgtcagtggcgtgctcctcgtcgtcctgcccagcagcatgatgactttgc 1191	
Db	83877	CCTCATGGTGTGTGTCAGTGGCGCTGCTCCGCTGCTCGCTGCCAGCAGCATGATGACTTTGC 83878	
QY	1192	tgatgacatctccctgctaaagtgaagagggcccatggcagagatagggattccctgg 1251	
Db	83817	TGATGACATCTCCCTGCTGAAGTGAAGAGGCCCATGGCAGAAAGATAGAGATTCCCTCG 83758	
QY	1252	accacacctcgtggttccatttgggtccacaaagtagagacacagatggcacctgtggcca 1311	
Db	83757	ACCACACCTCCGTTGTTCACTTTGGTCACAAGTAGGAGACACAGATGGCACCTGTGGCCA 83698	
QY	1312	gagcactcaggacctcccccacccacccaaatgcctctgcttgatggagaagaaagg 1371	
Db	83697	GAGCAGCTCAGGACCCCTCCCCACCCACCAATGCCTCTGCTGTGATGGAGAGAAAGG 83638	
QY	1372	ctggcaagtgttcaggagactgtacctgttagagacagaaagagaagaagca 1431	
Db	83637	CTGGCAAGTGTGGTTCAGGGACTGTACCTGTAGGAACAGAAAGAGAAAGAAAGCA 83578	
QY	1432	ctctgtgggggaatactcttgggtcacctcaaatattagtcggggaattctgctgttg 1491	
Db	83577	CTCTGTGGGGGAATACTCTTGGTCACCTCAAAATTTAAGTCGGGAATTTCTGCTGTG 83518	
QY	1492	aaacttcagccctgaaccttggc-caccattcctttaaatctccaccccaagattctt 1550	
Db	83517	AAACTTCAGCCCTGAACCTTTGTCCACCATTCCTTTAAATTCCTCAACCCCAAGTATTC 83458	
QY	1551	tctttcttagtttcagaagtactggcatcacacagcagggttaacttggcgtgtgtccctg 1610	
Db	83457	TCCTTTCTTAGTTTCAGAACTATGTCATCAGCAGGATTACTTGGCGTGTGTCCTTG 83398	
QY	1611	tggtaacctggcagaagagacaaagctgttttccctgtggtggccaaagtacgtaggaga 1670	
Db	83397	TGGTACCCAGGACAGAAAGACCAAGCTGTGTTCCTGCTGGCCAAAGTCAGTAGGAGA 83338	
QY	1671	ggatgcacgtttgctatttggcttttagagacagggactgtataaacagcctaacattgg 1730	
Db	83337	GGATGCACGTTTGTCTATTGCTTTTAGAGACAGGGACTGTATAAACAGCCTAACATTGG 83278	
QY	1731	tgcaaa 1736	
Db	83277	TGCAAA 83272	
RESULT	5		
LOCUS	AP000685	108772 bp DNA	
DEFINITION	Homo sapiens chromosome 11 clone CMB9-8M6 map 11q23, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	HTG 09-NOV-1999	

AP000685.1 GI:6277513  
HTG: HTGS\_PHASE1.  
Homo sapiens DNA, clone: CMB9-8M6.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 108772)  
Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
Homo sapiens 108,772 genomic DNA of 11q23  
Published Only in Database (1999) In press  
2 (bases 1 to 108772)  
Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
Direct Submission  
Submitted (08-NOV-1999) to the DDBJ/EMBL/GenBank databases.  
Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan  
(E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923, Fax: 81-42-778-9924)  
NOTE: This is a 'working draft' sequence. It currently consists of 27 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved  
1 10546 contig of 10546 of in length  
11047 17615 contig of 6569 of in length  
24086 contig of 5971 of in length  
24587 29736 contig of 5150 of in length  
30237 38079 contig of 7843 of in length  
38580 43515 contig of 4936 of in length  
44016 48745 contig of 4730 of in length  
49246 53614 contig of 4369 of in length  
54115 59119 contig of 5005 of in length  
59620 63763 contig of 4144 of in length  
64264 67186 contig of 2923 of in length  
67687 71093 contig of 3407 of in length  
71594 74598 contig of 3005 of in length  
75099 78391 contig of 3293 of in length  
78892 81382 contig of 2491 of in length  
81883 83560 contig of 1678 of in length  
84061 86317 contig of 2257 of in length  
86818 88559 contig of 1742 of in length  
89060 91079 contig of 2020 of in length  
91580 94007 contig of 2428 of in length  
94508 96680 contig of 2173 of in length  
97181 99237 contig of 2057 of in length  
99738 101029 contig of 1292 of in length  
101530 102776 contig of 1247 of in length  
103277 104568 contig of 1292 of in length  
105069 107111 contig of 2043 of in length  
107612 108772 contig of 1161 of in length.  
\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will be preserved.  
Location/Qualifiers  
1. .108772  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/clone="CMB9-8M6"  
/map="11q23"  
BASE COUNT 25040 a 21557 c 20852 g 26525 t 14798 others  
Query Match 20.0%; Score 350; DB 32; Length 108772;  
Best Local Similarity 99.2%; Pred. No. 3.3e-193;



AF190727 2158 bp mRNA ROD 26-OCT-1999  
Rattus norvegicus beta-site APP cleaving enzyme (Bace) mRNA,  
complete cds.  
AF190727  
AF190727.1 GI:6118542  
Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 2158)  
 AUTHORS Vassar, R., Bennett, B.D., Babu-Khan, S., Kahn, S., Mendiaz, E.A., Denis, P., Teplow, D.B., Ross, S., Amarante, P., Loeloff, R., Luo, Y., Fisher, S., Fuller, J., Edenson, S., Lile, J., Jarosinski, M.A., Biere, A.L., Curran, E., Burgess, T., Louis, J.-C., Collins, F., Treanor, J., Rogers, G. and Citron, M.

TITLE Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE

JOURNAL Science 286 (5440), 735-741 (1999)  
 MEDLINE 20002972

REFERENCE 2 (bases 1 to 2158)  
 AUTHORS Bennett, B.D., Vassar, R. and Citron, M.

TITLE Direct Submission

JOURNAL Submitted (29-SEP-1999) Neuroscience, Amgen Inc., One Amgen Center Dr., Thousand Oaks, CA 91320-1799, USA

FEATURES  
 source Location/Qualifiers  
 1. .2158 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 gene 1. .2158 /gene="Bace"  
 CDS 428..1933 /gene="Bace"  
 /codon\_start=1  
 /product="beta-site APP cleaving enzyme"  
 /protein\_id="AAF04144.1"  
 /db\_xref="GI:6118543"  
 /translation="MAPALRWLLLVGSGMLPAQGTGLGIRLPLRSLGAPPLGLRLP  
 RETDEPEPGRGSGFVEMVDNLKSGGQYVEMTVGGSPPTNLILVDTGSSNFAVG  
 AAPFLHRYQRLSSYRDRLKSVYPTQGWEGELGTDLVSIHPGPNVTVRANI  
 AATSDKFFINGSWEGILGLAYAEIARDPDSLEPFLSLVQTHIPNIFSLQCGA  
 GFLNQTALASVGGSMIIGGDHSLYTGSLWYTPIRREWYEVIVRVEINGQDLKM  
 DCKEYNDSIVDSGTNLRPKKFAEAVKSIKAASSTEKPDGFWLGEQLVCWQAG  
 TTPWNIFVLSYLMGEVNTQSFRTILPQOYLRPVEDVATSDQDCYKFAVSQSTGT  
 VMGAVIMEGFYVDFRKRIGFVACHVDFEFTAAVEGPFVTAADMDCGYNIPQT  
 DESTLMTIAYVMAICALEFMLPLCLMVQWRCRLCRHQHDDFADDSILLK"

BASE COUNT 458 a 640 c 619 g 441 t  
 ORIGIN

Query Match 3.0%; Score 53; DB 12; Length 2158;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-19;  
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 943 cgaattgcttgcgcagcgttcgcatgtgcacatgagttcagcagcgc 995  
 |||||  
 Db 1661 CGAATTGGCTTGTCTGCTGAGCGTTCGCGATGTCAGGATGAGTTCAGGAGCGC 1713

RESULT 8  
 AF190726  
 LOCUS AF190726 4001 bp mRNA ROD 26-OCT-1999  
 DEFINITION Mus musculus beta-site APP cleaving enzyme (Bace) mRNA, complete cds.  
 AUTHORS AF190726  
 VERSION AF190726.1 GI:6118540  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 4001)  
 AUTHORS Vassar, R., Bennett, B.D., Babu-Khan, S., Kahn, S., Mendiaz, E.A., Denis, P., Teplow, D.B., Ross, S., Amarante, P., Loeloff, R., Luo, Y., Fisher, S., Fuller, J., Edenson, S., Lile, J., Jarosinski, M.A., Biere, A.L., Curran, E., Burgess, T., Louis, J.-C., Collins, F., Treanor, J., Rogers, G. and Citron, M.

TITLE Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE

JOURNAL Science 286 (5440), 735-741 (1999)  
 MEDLINE 20002972

REFERENCE 2 (bases 1 to 4001)  
 AUTHORS Bennett, B.D., Vassar, R. and Citron, M.

TITLE Direct Submission  
 JOURNAL Submitted (29-SEP-1999) Neuroscience, Amgen Inc., One Amgen Center Dr., Thousand Oaks, CA 91320-1799, USA

FEATURES  
 source Location/Qualifiers  
 1. .4001 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 gene 1. .4001 /gene="Bace"  
 CDS 439..1934 /gene="Bace"  
 /codon\_start=1  
 /product="beta-site APP cleaving enzyme"  
 /protein\_id="AAF04143.1"  
 /db\_xref="GI:6118541"  
 /translation="MAPALRWLLLVGSGMLPAQGTGLGIRLPLRSLGAPPLGLRLP  
 RETDESEEPGRGSGFVEMVDNLKSGGQYVEMTVGGSPPTNLILVDTGSSNFAVG  
 AAPFLHRYQRLSSYRDRLKSVYPTQGWEGELGTDLVSIHPGPNVTVRANI  
 AATSDKFFINGSWEGILGLAYAEIARDPDSLEPFLSLVQTHIPNIFSLQCGA  
 GFLNQTALASVGGSMIIGGDHSLYTGSLWYTPIRREWYEVIVRVEINGQDLKM  
 DCKEYNDSIVDSGTNLRPKKFAEAVKSIKAASSTEKPDGFWLGEQLVCWQAG  
 TTPWNIFVLSYLMGEVNTQSFRTILPQOYLRPVEDVATSDQDCYKFAVSQSTGT  
 VMGAVIMEGFYVDFRKRIGFVACHVDFEFTAAVEGPFVTAADMDCGYNIPQT  
 DESTLMTIAYVMAICALEFMLPLCLMVQWRCRLCRHQHDDFADDSILLK"

BASE COUNT 921 a 1090 c 1055 g 935 t  
 ORIGIN

Query Match 3.0%; Score 53; DB 12; Length 4001;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-19;  
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 943 cgaattgcttgcgcagcgttcgcatgtgcacatgagttcagcagcgc 995  
 |||||  
 Db 1662 CGAATTGGCTTGTCTGCTGAGCGTTCGCGATGTCAGGATGAGTTCAGGAGCGC 1714

RESULT 9  
 AF196972/c  
 LOCUS AF196972 122568 bp DNA PRI 09-NOV-1999  
 DEFINITION Homo sapiens SSX4 protein gene, partial cds; OATL1 pseudogene, complete sequence; JM24 protein gene, partial cds; JM23 protein gene, complete cds; MG61 protein gene, partial cds; and phenylalkylamine binding protein gene, complete cds.  
 AUTHORS AF196972  
 VERSION AF196972.1 GI:6289067  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 122568)  
 AUTHORS Blechschmidt, X., Nyakatura, G., Strom, T.M., Drescher, B., Menzel, U., Meindl, A. and Rosenthal, A.

TITLE Direct Submission

JOURNAL Submitted (14-OCT-1999) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany

COMMENT 1-31666: cosmid J2332; 25789-61204: cosmid K0616; 57338-97932: cosmid I133; 82980-122568: cosmid A0842.

FEATURES  
 source Location/Qualifiers  
 1. .122568 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="X"  
 /map="Xp11.23"  
 /clone="cosmids J2332; K0616; I133; A0842"  
 complement(<1..5792)  
 gene /gene="SSX4"  
 complement(65..173)  
 repeat\_region /rpt\_family="MIR"  
 /evidence=not\_experimental  
 complement(join(<364..499,2486..2535,4399..4494,5171..5285,5724..5792))

CDS  
/gene="SSX4"  
/product="SSX4"  
complement(join(<364. .499,2486. .2535,4399. .4494,  
5171. .5285,5724. .5792))  
/gene="SSX4"  
/note="mRNA acc. no. U90841"  
/codon\_start=1  
/product="SSX4 protein"  
/protein\_id="RAF06796.1"  
/db\_xref="GI:6289068"  
translation="MNGDDAFARRPRDDAQLSEKLRKAFDDIAKYFSKKEWKMSSE  
KIVYVKLNLYEVTKLGKFKVTLPPFMESKRAADFGHNDGNDNRHNRQVRPQWTFG  
SLQRIFFKIMPKPAEENGLKEVPEASGPQNDGKQLCPGPNPSTLEKINKTS"  
complement(364. .499)  
/gene="SSX4"  
/number=5  
complement(1134. .1440)  
/rpt\_family="Alu"  
/evidence=not\_experimental  
repeat\_region  
complement(2038. .2166)  
/rpt\_family="MIR"  
/evidence=not\_experimental  
repeat\_region  
2199. .2313  
/rpt\_family="MIR"  
/evidence=not\_experimental  
exon  
complement(2486. .2535)  
/gene="SSX4"  
/number=4  
2866. .2917  
/rpt\_family="L1MB7"  
/evidence=not\_experimental  
repeat\_region  
3732. .4038  
/rpt\_family="AluSg"  
/evidence=not\_experimental  
exon  
complement(4399. .4494)  
/gene="SSX4"  
/number=3  
4537. .4832  
/rpt\_family="AluJo"  
/evidence=not\_experimental  
exon  
complement(5171. .5285)  
/gene="SSX4"  
/number=2  
complement(5724. .5792)  
/gene="SSX4"  
/number=1  
5862. .5926  
/rpt\_family="AluJ/FLAM"  
/evidence=not\_experimental  
repeat\_region  
complement(5927. .6106)  
/rpt\_family="AluSx"  
/evidence=not\_experimental  
repeat\_region  
complement(6904. .7345)  
/rpt\_family="L1PA2"  
/evidence=not\_experimental  
repeat\_region  
complement(7467. .7658)  
/rpt\_family="AluJb"  
/evidence=not\_experimental  
repeat\_region  
complement(7665. .7905)  
/rpt\_family="AluJo"  
/evidence=not\_experimental  
repeat\_region  
complement(7908. .7967)  
/rpt\_family="AluSg/x"  
/evidence=not\_experimental  
repeat\_region  
complement(7968. .8164)  
/rpt\_family="AluJb"  
/evidence=not\_experimental  
repeat\_region  
complement(8165. .8206)  
/rpt\_family="AluJo/FLAM"  
/evidence=not\_experimental  
repeat\_region  
8210. .8320  
/rpt\_family="AluSg/x"  
/evidence=not\_experimental

8786. .8867  
/rpt\_family="L1M4"  
/evidence=not\_experimental  
repeat\_region  
complement(9224. .9531)  
/rpt\_family="AluSg"  
/evidence=not\_experimental  
repeat\_region  
complement(10357. .10481)  
/rpt\_family="AluYb8"  
/evidence=not\_experimental  
repeat\_region  
complement(10487. .10965)  
/rpt\_family="L1MC/D"  
/evidence=not\_experimental  
repeat\_region  
complement(10966. .10983)  
/rpt\_family="AluSg"  
/evidence=not\_experimental  
repeat\_region  
complement(11025. .11317)  
/rpt\_family="AluSg"  
/evidence=not\_experimental  
repeat\_region  
complement(11318. .11927)  
/rpt\_family="L1MC/D"  
/evidence=not\_experimental  
repeat\_region  
complement(11932. .12223)  
/rpt\_family="AluSg"  
/evidence=not\_experimental  
misc\_feature  
12112. .12228  
/note="CpG island score = 0.91, GC = 63.20%, CpGs = 10"  
/note="Region: CpG island"  
/evidence=not\_experimental  
repeat\_region  
12359. .12701  
/rpt\_family="L1P"  
/evidence=not\_experimental  
repeat\_region  
12826. .12941  
/rpt\_family="HAL1"  
/evidence=not\_experimental  
repeat\_region  
12942. .13249  
/rpt\_family="AluSx"  
/evidence=not\_experimental  
repeat\_region  
13250. .13311  
/rpt\_family="HAL1"  
/evidence=not\_experimental  
repeat\_region  
13290. .13317  
/rpt\_family="L1MD2"  
/evidence=not\_experimental  
repeat\_region  
13318. .13572  
/rpt\_family="AluSg"  
/evidence=not\_experimental  
repeat\_region  
13573. .13827  
/rpt\_family="L1MD2"  
/evidence=not\_experimental  
repeat\_region  
14020. .14334  
/rpt\_family="L1MD2"  
/evidence=not\_experimental  
repeat\_region  
14335. .14655  
/rpt\_family="AluSx"  
/evidence=not\_experimental  
repeat\_region  
14656. .15156  
/rpt\_family="L1MD2"  
/evidence=not\_experimental  
repeat\_region  
15157. .15480  
/rpt\_family="AluY"  
/evidence=not\_experimental  
repeat\_region  
15481. .15734  
/rpt\_family="L1MD2"  
/evidence=not\_experimental  
repeat\_region  
complement(15735. .16036)  
/rpt\_family="AluSx"  
/evidence=not\_experimental  
gene  
complement(16068. .42262)  
/gene="OATL1"  
/note="ornithine aminotransferase-like 1"  
/pseudo  
16909. .16968  
/rpt\_family="AluS"



```

* 9233 9794: contig of 562 bp in length
* 9795 9874: gap of unknown length
* 9875 10534: contig of 660 bp in length
* 10535 10614: gap of unknown length
* 10615 11280: contig of 666 bp in length
* 11281 11360: gap of unknown length
* 11361 12508: contig of 1148 bp in length
* 12509 12588: gap of unknown length
* 12589 13169: contig of 581 bp in length
* 13170 13249: gap of unknown length
* 13250 13913: contig of 664 bp in length
* 13914 13993: gap of unknown length
* 13994 14881: contig of 888 bp in length
* 14882 14961: gap of unknown length
* 14962 15607: contig of 646 bp in length
* 15608 15687: gap of unknown length
* 15688 16540: contig of 853 bp in length
* 16541 16620: gap of unknown length
* 16621 17376: contig of 756 bp in length
* 17377 17456: gap of unknown length
* 17457 18185: contig of 729 bp in length
* 18186 18265: gap of unknown length
* 18266 19475: contig of 1210 bp in length
* 19476 19555: gap of unknown length
* 19556 19799: contig of 424 bp in length
* 19980 20059: gap of unknown length
* 20060 21183: contig of 1124 bp in length
* 21184 21263: gap of unknown length
* 21264 22130: contig of 867 bp in length
* 22131 22310: gap of unknown length
* 22311 23537: contig of 1327 bp in length
* 23538 23617: gap of unknown length
* 23618 24710: contig of 1093 bp in length
* 24711 24790: gap of unknown length
* 24791 25767: contig of 977 bp in length
* 25768 25847: gap of unknown length
* 25848 26816: contig of 989 bp in length
* 26817 26896: gap of unknown length
* 26897 27821: contig of 925 bp in length
* 27822 27901: gap of unknown length
* 27902 29209: contig of 1308 bp in length
* 29210 29289: gap of unknown length
* 29290 30136: contig of 847 bp in length
* 30137 30216: gap of unknown length
* 30217 31091: contig of 875 bp in length
* 31092 31171: gap of unknown length
* 31172 32461: contig of 1290 bp in length
* 32462 32541: gap of unknown length
* 32542 33440: contig of 899 bp in length
* 33441 33520: gap of unknown length
* 33521 34642: contig of 1122 bp in length
* 34643 34722: gap of unknown length
* 34723 35977: contig of 1255 bp in length
* 35978 36057: gap of unknown length
* 36058 36877: contig of 820 bp in length
* 36878 36957: gap of unknown length
* 36958 37339: contig of 982 bp in length
* 37340 38019: gap of unknown length
* 38020 39018: contig of 1019 bp in length
* 39019 39118: gap of unknown length
* 39119 39904: contig of 786 bp in length
* 39905 39984: gap of unknown length
* 39985 41091: contig of 1107 bp in length
* 41092 41171: gap of unknown length
* 41172 42035: contig of 864 bp in length
* 42036 42115: gap of unknown length
* 42116 43447: contig of 1332 bp in length
* 43448 43527: gap of unknown length
* 43528 44892: contig of 1365 bp in length
* 44893 44972: gap of unknown length
* 44973 45896: contig of 924 bp in length
* 45897 45976: gap of unknown length
* 45977 47085: contig of 1109 bp in length

```

```

* 47086 47165: gap of unknown length
* 47166 48337: contig of 1172 bp in length
* 48338 48417: gap of unknown length
* 48418 49953: contig of 1536 bp in length
* 49954 50033: gap of unknown length
* 50034 50906: contig of 873 bp in length
* 50907 50986: gap of unknown length
* 50987 52003: contig of 1017 bp in length
* 52004 52083: gap of unknown length
* 52084 52864: contig of 781 bp in length
* 52865 52944: gap of unknown length
* 52945 54405: contig of 1461 bp in length
* 54406 54485: gap of unknown length
* 54486 56164: contig of 1679 bp in length
* 56165 56244: gap of unknown length
* 56245 57586: contig of 1342 bp in length
* 57587 57666: gap of unknown length
* 57667 58566: contig of 900 bp in length
* 58567 58646: gap of unknown length
* 58647 60263: contig of 1617 bp in length
* 60264 60343: gap of unknown length
* 60344 61865: contig of 1522 bp in length
* 61866 61945: gap of unknown length
* 61946 64811: contig of 2866 bp in length
* 64812 64891: gap of unknown length
* 64892 66567: contig of 1676 bp in length
* 66568 66647: gap of unknown length
* 66648 68427: contig of 1780 bp in length
* 68428 68507: gap of unknown length
* 68508 71124: contig of 2617 bp in length
* 71125 71204: gap of unknown length
* 71205 73997: contig of 2793 bp in length
* 73998 74078: gap of unknown length
* 74079 75668: contig of 1591 bp in length
* 75669 75748: gap of unknown length
* 75749 79707: contig of 3959 bp in length
* 79708 79787: gap of unknown length
* 79788 83303: contig of 3516 bp in length
* 83304 83384: gap of unknown length
* 83385 85456: contig of 2073 bp in length
* 85457 85536: gap of unknown length
* 85537 88306: contig of 2770 bp in length
* 88307 88386: gap of unknown length
* 88387 91501: contig of 3115 bp in length
* 91502 91582: gap of unknown length
* 91583 93395: contig of 3734 bp in length
* 93396 95316: gap of unknown length
* 95317 100403: contig of 5008 bp in length
* 100404 100483: gap of unknown length
* 100484 103372: contig of 4889 bp in length
* 103373 103452: gap of unknown length
* 103453 106053: contig of 601 bp in length
* 106054 106133: gap of unknown length
* 106134 106538: contig of 405 bp in length

```

Query Match 1.3%; Score 22; DB 33; Length 121043;  
 Best Local Similarity 100.0%; Pred. No. 0.45;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1726 attggtgcaaaaaa 1747  
 |||||  
 DB 48072 ATTGTCACAAAAA 48093

RESULT 12  
 HSJ729N16  
 LOCUS HSJ729N16 154469 bp DNA  
 DEFINITION Human DNA sequence \*\*\* SEQUENCING IN PROGRESS \*\*\* from clone  
 dj729N16, complete sequence.  
 ACCESSION AL096707  
 VERSION AL096707.12 GI:5791542  
 KEYWORDS HTG.  
 SOURCE human.





---









```

Query Match          9.4%; Score 164; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 2.8e-63;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gctcaccagtgtcccttcattctatctgcaagccacattacactctgttctggtgg 60
   |||
Db 18 GCTCACCAGTGTCCCTTCATCTATCTCAAGCCACATTTACACTCTTCTGGGTGG 77
   |||

QY 61 tccagcacatccgggacccctccgaaggtgtgtatgtgccctacacccagggaagtgg 120
   |||
Db 78 TCCAGCACATACCGGGACCTCCGGAAGGGTGTGTATGTGCCCTACACCCAGGCAAGTGG 137
   |||

QY 121 gaagggagctggccagccactggtaagcatcccccatggccc 164
   |||
Db 138 GAAGGGAGCTGGCGACCGACCTGGTAAGCATCCCATGGCCCC 181
   |||

RESULT 4
ID X27311 standard; DNA; 1882 BP.
AC X27311;
DE Human secreted protein gene 1 clone HCEIA77.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
PN W0902546-AL.
PF 07-JUL-1998; U13684.
PR 12-SEP-1997; US-058785.
PR 08-JUL-1997; US-051916.
PR 08-JUL-1997; US-051918.
PR 08-JUL-1997; US-051919.
PR 08-JUL-1997; US-051920.
PR 08-JUL-1997; US-051925.
PR 08-JUL-1997; US-051926.
PR 08-JUL-1997; US-051928.
PR 08-JUL-1997; US-051929.
PR 08-JUL-1997; US-051930.
PR 08-JUL-1997; US-051931.
PR 08-JUL-1997; US-051932.
PR 08-JUL-1997; US-052732.
PR 08-JUL-1997; US-052733.
PR 08-JUL-1997; US-052793.
PR 08-JUL-1997; US-052795.
PR 08-JUL-1997; US-052803.
PR 18-AUG-1997; US-055684.
PR 18-AUG-1997; US-055722.
PR 18-AUG-1997; US-055723.
PR 18-AUG-1997; US-055947.
PR 18-AUG-1997; US-055948.
PR 18-AUG-1997; US-055949.
PR 18-AUG-1997; US-055950.
PR 18-AUG-1997; US-055953.
PR 18-AUG-1997; US-055954.
PR 18-AUG-1997; US-055964.
PR 18-AUG-1997; US-055984.
PR 18-AUG-1997; US-056360.
PR 12-SEP-1997; US-058660.
PR 12-SEP-1997; US-058661.
PR 12-SEP-1997; US-058664.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Ebner R, Fischer CL, Kyaw H, Lafleur DW, Li Y, Moore PA,
PI Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR, zeng Z;
DR WPI; 99-120770/10.
DR P-PSDB; Y02650.
PT New isolated human genes and the secreted polypeptides they encode -

```

```

PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1; Page 242; 464pp; English.
CC This sequence represents a nucleic acid molecule which encodes a secreted
CC human protein. The gene number, and the clone it is derived from, are
CC detailed in the descriptor line. The gene can be used to generate fusion
CC proteins by linking to the gene to a human immunoglobulin Fc portion
CC (e.g. X27302) for increasing the stability of the fused protein as
CC compared to the human protein only.
CC The invention relates to 123 novel genes and their fragments (nucleic
CC acid sequences: X27311-X27449; amino acid sequences Y02650-Y02788) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 123 polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of: cancer; tumours;
CC neurodegenerative disorders; developmental abnormalities and foetal
CC deficiencies; blood disorders; leukaemias; diseases of the immune system
CC (including allergies or asthma); lymphocytic diseases; brain associated
CC diseases; hepatic and renal disease; lymphomas; inflammation; ischaemic
CC shock; Alzheimer's and cognitive disorders; schizophrenia; restenosis;
CC prostate diseases; obesity; disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies; diseases of testis, lung or
CC thymus; thyroiditis and thyroid tumours; digestive/endocrine disorders,
CC including metabolic regulation, malabsorption, gastritis and neoplasms;
CC and AIDS.
SQ Sequence 1882 BP; 507 A; 445 C; 442 G; 479 T;

Query Match          1.2%; Score 21; DB 1; Length 1882;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1727 ttgggtgcaaaaaaataaaaaa 1747
   |||
Db 1857 TTGGTGCACAAAAAATAAAAAA 1877
   |||

RESULT 5
ID V75741 standard; DNA; 115 BP.
AC V75741;
DE 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #1430.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
PN EP-786519-A2.
PF 30-JUL-1997.
PR 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI; 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 1954; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of

```

```

FT      CDS          179..763
FT      FT           /*tag= a
FT
FT      WO9747740-A2.
FT      PD           18-DEC-1997.
FT      PF           11-JUN-1997; G01573
FT      PR           11-JUN-1996; GB-012214.
FT      PA           (MLCW ) MALLINCKRODT VETERINARY INC.
FT      PI           Andrews SJ, Dalton JP;
FT      PP           WPI; 98-052307/05.
FT      DR           P-PADB; W46775.
FT      PT           Vaccine for controlling infection by helminth parasites - containing
FT      PT           new peroxiredoxin or beta-tubulin, particularly used to combat liver
FT      PT           fluke
FT      PS           Claim 11; Fig 4; 4lpp; English.
FT      CC           The present sequence encodes a peroxiredoxin antigenic fragment. The
FT      CC           sequence was isolated from a cDNA library of Fasciola hepatica. The
FT      CC           protein can be used in a vaccine for controlling parasitic helminth
FT      CC           infections in mammals (beta-tubulin antigenic fragments can also be
FT      CC           used, see V16368). The vaccines are used to treat cattle and sheep
FT      CC           but may also be used in humans, pets etc. Attenuated or virulent
FT      CC           viruses, or host cells, containing the nucleic acid may be used as
FT      CC           live or inactivated vaccines.
FT      SQ           Sequence 1163 BP; 283 A; 297 C; 295 G; 278 T;

Query Match          1.1%; Score 19; DB 1; Length 1163;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1729 ggtgcataaaaaaa 1747
Db    828 GGTGCAAAAAA 846

RESULT 8
T36790 ID T36790 standard; DNA; 1242 BP.
AC T36790;
DC 04-NOV-1996 (first entry)
DE Prostate-specific membrane antigen gene intron 2 (reverse).
KW Prostate-specific membrane antigen; PSM; prostate cancer;
KW metastasis; gene therapy; diagnosis; ds.
OS Homo sapiens.
FH Key
FT misc_difference 381..388
FT FT           /*tag= a
FT FT           /note= "n bases at positions 381-388 are not
FT FT           identified in the specification"
FT FT           misc_difference 676..680
FT FT           /*tag= b
FT FT           /note= "n bases at positions 676-680 are not
FT FT           identified in the specification"
FT FT           misc_difference 721
FT FT           /*tag= c
FT FT           /note= "base n at position 721 is not identified
FT FT           in the specification"
FT FT           misc_difference 975
FT FT           /*tag= d
FT FT           /note= "base n at position 975 is not identified
FT FT           in the specification"
FT FT           misc_difference 979
FT FT           /*tag= e
FT FT           /note= "base n at position 979 is not
FT FT           identified in the specification"
FT FT           misc_difference 982
FT FT           /*tag= f
FT FT           /note= "base n at position 982 is not
FT FT           identified in the specification"
FT FT           misc_difference 1091
FT FT           /*tag= g
FT FT           /note= "base n at position 1091 is not identified
FT FT           in the specification"
FT FT           misc_difference 1167..1172

```

```

FT      /*tag= h
FT      /note= "n bases at positions 1167-1172 are not
FT      identified in the specification"
FT      misc_difference 1212
FT      /*tag= i
FT      /note= "base n at position 1212 is not identified
FT      in the specification"
FT      WO9626272-A1.
PN      29-AUG-1996.
PD      23-FEB-1996; US-394152.
PR      24-FEB-1995; US-470735.
PR      02-JUN-1995; US-470735.
PR      02-JUN-1995; US-466381.
PA      (SLOK ) SLOAN KETTERING INST CANCER RES.
PI      Fair WR, Heston WDW, Israeli RS;
PI      WPI; 96-402365/40.
PT      DNA encoding alternatively spliced prostate-specific membrane
PT      antigen - useful to develop prods. for detecting haematogenous
PT      micrometastatic tumour cells, or prostate cancer progression
PT      Disclosures: Fig 75A-C; 284pp; English.
PS      Sequences are provided for introns 1F, 1R, 2F, 2R, 3F, 3R, 4F and
CC      4R (T36787-94, respectively) of the human gene (see also T36785
CC      for cDNA sequence) coding for prostate-specific membrane (PSM) antigen
CC      (W022334).
CC      (NB. Different sequences for introns 1-4 (forward and reverse) are
CC      given in T36787-94).
CC      489 A; 198 C; 228 G; 302 T;
SQ      Sequence 1242 BP;
Query Match 1.1%; Score 19; DB 1; Length 1242;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1729 ggtgcataaaataaaaaa 1747
Db 1193 GGTGCATAAAATAAAAAA 1211
RESULT 9
V23931/C
ID V23931 standard; DNA; 1257 BP.
AC V23931;
DE 03-AUG-1998 (first entry)
DE Marker DNA from blight-suffering plant.
DE KW Blight-suffering plant; marker DNA; blight-resistance; ds.
OS Nicotiana tabacum.
PN J10127290-A.
PD 13-MAY-1998.
PF 31-OCT-1996; 290847.
PR 31-OCT-1996; JP-290847.
PA (NISE ) JAPAN TOBACCO INC.
WPI; 98-340666/30.
DR Selection of blight-resistant plant - by identification of DNA
PT selection marker present in blight suffering plant
PS Claim 8; Page 10; 16pp; Japanese.
CC This sequence represents a marker DNA from Nicotiana tabacum, and is used
CC in the method of the invention. The method is for selecting a
CC blight-resistant plant by comprises identification of a specific DNA
CC marker present in a blight-suffering plant and absent in a
CC blight-resistant plant. The method can select a resistant plant easily.
CC Sequence 1257 BP; 310 A; 319 C; 265 G; 363 T;
SQ
Query Match 1.1%; Score 19; DB 1; Length 1257;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1410 cagaaagagagaagaa 1428
Db 788 CAGAAAGAGAGAAAGAA 770
RESULT 10

```









**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 10:58:55 ; Search time 181.17 Seconds  
(without alignments)  
1154.476 Million cell updates/sec

Title: US-09-215-435-66  
Perfect score: 1747  
Sequence: 1 gctccacaatggtccctt.....tggtcacaaaaaaaaa 1747

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 214294 seqs, 59861574 residues

Word size : 0

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2.6/ptodata/1/ina/5C\_COMB.seq.\*  
4: /cgn2.6/ptodata/1/ina/5D\_COMB.seq.\*  
5: /cgn2.6/ptodata/1/ina/6\_COMB.seq.\*  
6: /cgn2.6/ptodata/1/ina/PTUS9\_COMB.seq.\*  
7: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	1.1	4698	1 US-07-807-043B-5	Sequence 5, Appl
C 2	19	1.1	4698	1 US-08-299-849B-5	Sequence 5, Appl
C 3	19	1.1	4698	4 US-08-142-368A-5	Sequence 5, Appl
4	18	1.0	651	1 US-08-171-385-4	Sequence 4, Appl
5	18	1.0	1472	1 US-08-333-565-1	Sequence 1, Appl
6	18	1.0	1472	3 US-08-661-479-1	Sequence 1, Appl
7	18	1.0	1944	7 5260223-2	Patent No. 5260223
C 8	18	1.0	12912	3 US-08-460-751-1	Sequence 1, Appl
C 9	18	1.0	31571	1 US-08-323-443B-1	Sequence 1, Appl
C 10	17	1.0	20	4 US-08-715-461-3	Sequence 3, Appl
C 11	17	1.0	20	4 US-08-715-461-4	Sequence 4, Appl
12	17	1.0	57	1 US-08-120-827-93	Sequence 93, Appl
13	17	1.0	57	2 US-08-478-675-93	Sequence 93, Appl
14	17	1.0	83	2 US-08-120-827-100	Sequence 100, App
15	17	1.0	83	2 US-08-478-675-100	Sequence 100, App
16	17	1.0	177	3 US-09-032-684-20	Sequence 20, Appl
17	17	1.0	279	2 US-08-686-878A-50	Sequence 50, Appl
18	17	1.0	279	2 US-08-721-489-4	Sequence 4, Appl
19	17	1.0	306	2 US-08-627-706-8	Sequence 8, Appl
20	17	1.0	500	2 US-08-627-706-9	Sequence 9, Appl
21	17	1.0	539	1 US-08-370-648-18	Sequence 18, Appl
22	17	1.0	539	3 US-08-531-662B-18	Sequence 18, Appl
23	17	1.0	540	1 US-08-370-648-16	Sequence 16, Appl
24	17	1.0	540	3 US-08-531-662B-16	Sequence 16, Appl
25	17	1.0	588	5 US-08-744-138-1	Sequence 1, Appl
26	17	1.0	600	4 US-08-187-186A-1	Sequence 1, Appl

27	17	1.0	600	4 US-08-442-497C-1	Sequence 1, Appl
28	17	1.0	600	6 PCT-US94-05186-1	Sequence 1, Appl
c 29	17	1.0	645	1 US-08-329-055-1	Sequence 1, Appl
30	17	1.0	686	3 US-08-522-421-7	Sequence 7, Appl
31	17	1.0	690	5 US-08-865-297-3	Sequence 3, Appl
32	17	1.0	770	5 US-08-865-297-5	Sequence 3, Appl
33	17	1.0	833	2 US-08-450-944-1	Sequence 5, Appl
34	17	1.0	833	6 PCT-US96-07709-1	Sequence 1, Appl
c 35	17	1.0	833	6 PCT-US96-07709-13	Sequence 1, Appl
36	17	1.0	857	1 US-08-308-883-1	Sequence 13, Appl
37	17	1.0	857	2 US-08-730-163-1	Sequence 1, Appl
38	17	1.0	862	1 US-08-379-280-4	Sequence 1, Appl
39	17	1.0	870	2 US-08-203-747-7	Sequence 4, Appl
40	17	1.0	870	2 US-08-458-298-7	Sequence 7, Appl
41	17	1.0	874	2 US-08-469-667-3	Sequence 7, Appl
42	17	1.0	874	6 PCT-US95-07289-3	Sequence 3, Appl
43	17	1.0	885	5 US-08-865-297-1	Sequence 3, Appl
44	17	1.0	1045	3 US-08-808-550-1	Sequence 1, Appl
45	17	1.0	1046	3 US-08-808-550-40	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-07-807-043B-5/c  
; Sequence 5, Application US/07807043B  
; Patent No. 5342774  
; GENERAL INFORMATION:  
; APPLICANT: Boon, Thierry, Van den Eynde, Beno t  
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor  
; TITLE OF INVENTION: Rejection Antigens and Uses Thereof  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felfe & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York City  
; STATE: New York  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
; COMPUTER: IBM  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/807,043B  
; FILING DATE: 19911212  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/764,364  
; FILING DATE: 23-SEPTEMBER-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/728,838  
; FILING DATE: 9-JULY-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/705,702  
; FILING DATE: 23-May-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hanson, No. 5342774man D.  
; REGISTRATION NUMBER: 30,946  
; REFERENCE/DOCKET NUMBER: LUD 253.3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 688-9200  
; TELEFAX: (212) 838-3884  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4698 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: singular  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
US-07-807-043B-5

Query Match 1.1%; Score 19; DB 1; Length 4698;  
Best Local Similarity 100.0%; Pred. No. 7.4;

Qy	1729 ggtgcacccccccccc	1747
	{cccccccccccccccc}	
Db	2031 GGTCACACACACACAAA	2013

```

; RESULT 5
; US-08-333-565-1
; Sequence 1, Application US/08333565
; Patent No. 5622852
; GENERAL INFORMATION:
; APPLICANT: KORSMEYER, Stanley J.
; TITLE OF INVENTION: Bcl-x/Bcl-2 ASSOCIATED CELL DEATH
; TITLE OF INVENTION: REGULATOR
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: US
; ZIP: 94301
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/333,565
; FILING DATE: 31-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M
; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 15726A-000700
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1472 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-333-565-1

Query Match 1.0%; Score 18; DB 1; Length 1472;
Best Local Similarity 100.0%; Pred.No.22;
Matches 18; Conservative 0; Mismatches 0; Indels

QY 1730 gtgcacaaataaaaaaaa 1747
      |||iiiiiiiii|||
Db 1451 GTGCAAAAAAAAAAAAAA 1468

RESULT 6
US-08-661-479-1
; Sequence 1, Application US/08661479
; Patent No. 5834209
; GENERAL INFORMATION:
; APPLICANT: KORSMEYER, Stanley J.
; TITLE OF INVENTION: Bcl-x/Bcl-2 ASSOCIATED CELL DEATH
; TITLE OF INVENTION: REGULATOR
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Khourie and Crew
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: US
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/661,479
; FILING DATE: 11-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/333,565
; FILING DATE: 31-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M
; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 15726A-000700
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1472 base pairs
; TYPE: nucleic acid

```



;  
; LENGTH: 31571 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: PKD1 GENOMIC  
; US-08-323-443B-1

Query Match 1.0%; Score 18; DB 1; Length 31571;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 523 caggatctgaaatggac 540  
|||||

DB 28229 CAGGATCTGAAATGGAC 28212

## RESULT 10

US-08-715-461-3/c  
; Sequence 3, Application US/08715461

; Patent No. 5985556

; GENERAL INFORMATION:

; APPLICANT: KAMBARA, Hideki

; APPLICANT: OKANO, Kazunori

; TITLE OF INVENTION: DNA SEQUENCING METHOD AND DNA SAMPLE

; TITLE OF INVENTION: PREPARATION METHOD

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ANTONELLI, TERRY STOUT & KRAUS

; STREET: 1300 No. 5985556th Seventeenth Street, Suite 1800

; CITY: Arlington

; STATE: VA

; COUNTRY: USA

; ZIP: 22209

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/715,461

; FILING DATE: 18-SEP-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: TERRY, David T.

; REGISTRATION NUMBER: 20,178

; REFERENCE/DOCKET NUMBER: 500.34872X00

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-312-6600

; TELEFAX: 703-312-6666

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 20 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: other nucleic acid

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-715-461-3

## Query Match

1.0%; Score 17; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 71;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1731 tgcaaaaaa 1747

|||||

DB 17 TGCAAAAAAAAAAAAA 1

## RESULT 11

US-08-715-461-4/c

; Sequence 4, Application US/08715461

; Patent No. 5985556

; GENERAL INFORMATION:

; APPLICANT: KAMBARA, Hideki

; APPLICANT: OKANO, Kazunori

; TITLE OF INVENTION: DNA SEQUENCING METHOD AND DNA SAMPLE

; TITLE OF INVENTION: PREPARATION METHOD

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ANTONELLI, TERRY STOUT & KRAUS

; STREET: 1300 No. 5985556th Seventeenth Street, Suite 1800

; CITY: Arlington

; STATE: VA

; COUNTRY: USA

; ZIP: 22209

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/715,461

; FILING DATE: 18-SEP-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: TERRY, David T.

; REGISTRATION NUMBER: 20,178

; REFERENCE/DOCKET NUMBER: 500.34872X00

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-312-6600

; TELEFAX: 703-312-6666

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 20 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: other nucleic acid

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-715-461-4

## Query Match

1.0%; Score 17; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 71;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1731 tgcaaaaaa 1747

|||||

DB 17 TGCAAAAAAAAAAAAA 1

## RESULT 12

US-08-120-827-93

; Sequence 93, Application US/08120827

; Patent No. 5523495

; GENERAL INFORMATION:

; APPLICANT: KEENE, JACK D.

; APPLICANT: KING, PETER H.

; APPLICANT: LEVINE, TODD

; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE

; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS

; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION

; NUMBER OF SEQUENCES: 101

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

; ADDRESSEE: P.C.

; STREET: 1755 Jefferson Davis Highway, Fourth Floor

RESULT 13  
US-08-478-675-93  
: Sequence 93, Application US/08478675  
: Patent No. 5773246  
: GENERAL INFORMATION:  
: APPLICANT: KEENE, JACK D.  
: APPLICANT: KING, PETER H.  
: APPLICANT: LEVINE, TODD  
: TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE  
: TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF  
: TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND  
: NUMBER OF SEQUENCES: 101  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: OBOLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
: ADDRESSEE: P.C.  
: STREET: 1755 Jefferson Davis Highway, Fourth Floor  
: CITY: Arlington  
: STATE: Virginia  
: COUNTRY: U.S.A.  
: ZIP: 22202  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/478,675  
: FILING DATE: 07-JUN-1996  
: CLASSIFICATION: 536  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/120,827  
: FILING DATE: 15-SEP-1993  
: ATTORNEY/AGENT INFORMATION:

```

// INFORMATION FOR SEQ ID NO: 100
//
// SEQUENCE CHARACTERISTICS:
//     LENGTH: 83 base pairs
//     TYPE: nucleic acid
//     STRANDEDNESS: unknown
//     TOPOLOGY: unknown
//     MOLECULE TYPE: cDNA
US-08-120-827-100

```

**RESULT 14**



Query Match 1.0%; Score 17; DB 1; Length 83;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1731 tgcaaaaaaaaaaaaa 1747  
Db 62 TGCACAAAAAAAAAAAAA 78

RESULT 15  
US-08-478-675-100  
; Sequence 100; Application US/08478675  
; Patent No. 5773246  
; GENERAL INFORMATION:  
; APPLICANT: KEENE, JACK D.  
; APPLICANT: KING, PETER H.  
; APPLICANT: LEVINE, TODD  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE  
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS  
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION  
; NUMBER OF SEQUENCES: 101  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 Jefferson Davis Highway, Fourth Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478,675  
; FILING DATE: 07-JUN-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/120,827  
; FILING DATE: 15-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 5773246man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 714-158-0 CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)413-3000  
; TELEFAX: (703)413-2220  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 100:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 83 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cdna  
; US-08-478-675-100

Query Match 1.0%; Score 17; DB 2; Length 83;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1731 tgcaaaaaaaaaaaaa 1747  
Db 62 TGCACAAAAAAAAAAAAA 78

Search completed: May 1, 2000, 16:01:51  
Job time: 18176 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 13:53:36 ; Search time 4088.29 Seconds  
(without alignments)  
-1465.366 Million cell updates/sec

Title: US-09-215-435-66  
Perfect score: 1747  
Sequence: 1 gctcaccatgttccctt.....tggtycaaaaaaaaaaaaaa 1747

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 5265378 seqs, -1714608768 residues

Word size : 0

Total number of hits satisfying chosen parameters: 10530736

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Pending\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/pna/PCTUS9\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US081A\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US081B\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US081C\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US081D\_COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US082A\_COMB.seq.\*  
9: /cgn2\_6/ptodata/1/pna/US082B\_COMB.seq.\*  
10: /cgn2\_6/ptodata/1/pna/US082C\_COMB.seq.\*  
11: /cgn2\_6/ptodata/1/pna/US083A\_COMB.seq.\*  
12: /cgn2\_6/ptodata/1/pna/US083B\_COMB.seq.\*  
13: /cgn2\_6/ptodata/1/pna/US084A\_COMB.seq.\*  
14: /cgn2\_6/ptodata/1/pna/US084C\_COMB.seq.\*  
15: /cgn2\_6/ptodata/1/pna/US084C\_COMB.seq.\*  
16: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*  
17: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*  
18: /cgn2\_6/ptodata/1/pna/US087A\_COMB.seq.\*  
19: /cgn2\_6/ptodata/1/pna/US087B\_COMB.seq.\*  
20: /cgn2\_6/ptodata/1/pna/US087C\_COMB.seq.\*  
21: /cgn2\_6/ptodata/1/pna/US088A\_COMB.seq.\*  
22: /cgn2\_6/ptodata/1/pna/US088B\_COMB.seq.\*  
23: /cgn2\_6/ptodata/1/pna/US088C\_COMB.seq.\*  
24: /cgn2\_6/ptodata/1/pna/US088D\_COMB.seq.\*  
25: /cgn2\_6/ptodata/1/pna/US089A\_COMB.seq.\*  
26: /cgn2\_6/ptodata/1/pna/US089B\_COMB.seq.\*  
27: /cgn2\_6/ptodata/1/pna/US089C\_COMB.seq.\*  
28: /cgn2\_6/ptodata/1/pna/US089D\_COMB.seq.\*  
29: /cgn2\_6/ptodata/1/pna/US089E\_COMB.seq.\*  
30: /cgn2\_6/ptodata/1/pna/US089F\_COMB.seq.\*  
31: /cgn2\_6/ptodata/1/pna/US089G\_COMB.seq.\*  
32: /cgn2\_6/ptodata/1/pna/US089H\_COMB.seq.\*  
33: /cgn2\_6/ptodata/1/pna/US089I\_COMB.seq.\*  
34: /cgn2\_6/ptodata/1/pna/US090A\_COMB.seq.\*  
35: /cgn2\_6/ptodata/1/pna/US090B\_COMB.seq.\*  
36: /cgn2\_6/ptodata/1/pna/US090C\_COMB.seq.\*  
37: /cgn2\_6/ptodata/1/pna/US090D\_COMB.seq.\*  
38: /cgn2\_6/ptodata/1/pna/US090E\_COMB.seq.\*  
39: /cgn2\_6/ptodata/1/pna/US091A\_COMB.seq.\*  
40: /cgn2\_6/ptodata/1/pna/US091B\_COMB.seq.\*  
41: /cgn2\_6/ptodata/1/pna/US091C\_COMB.seq.\*  
42: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq.\*  
43: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq.\*

44: /cgn2\_6/ptodata/1/pna/US092C\_COMB.seq.\*  
45: /cgn2\_6/ptodata/1/pna/US092D\_COMB.seq.\*  
46: /cgn2\_6/ptodata/1/pna/US092E\_COMB.seq.\*  
47: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq.\*  
48: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq.\*  
49: /cgn2\_6/ptodata/1/pna/US093C\_COMB.seq.\*  
50: /cgn2\_6/ptodata/1/pna/US093D\_COMB.seq.\*  
51: /cgn2\_6/ptodata/1/pna/US093E\_COMB.seq.\*  
52: /cgn2\_6/ptodata/1/pna/US093F\_COMB.seq.\*  
53: /cgn2\_6/ptodata/1/pna/US094A\_COMB.seq.\*  
54: /cgn2\_6/ptodata/1/pna/US094B\_COMB.seq.\*  
55: /cgn2\_6/ptodata/1/pna/US094C\_COMB.seq.\*  
56: /cgn2\_6/ptodata/1/pna/US094D\_COMB.seq.\*  
57: /cgn2\_6/ptodata/1/pna/US094E\_COMB.seq.\*  
58: /cgn2\_6/ptodata/1/pna/US094F\_COMB.seq.\*  
59: /cgn2\_6/ptodata/1/pna/US094G\_COMB.seq.\*  
60: /cgn2\_6/ptodata/1/pna/US094H\_COMB.seq.\*  
61: /cgn2\_6/ptodata/1/pna/US094I\_COMB.seq.\*  
62: /cgn2\_6/ptodata/1/pna/US094J\_COMB.seq.\*  
63: /cgn2\_6/ptodata/1/pna/US094K\_COMB.seq.\*  
64: /cgn2\_6/ptodata/1/pna/US094L\_COMB.seq.\*  
65: /cgn2\_6/ptodata/1/pna/US094M\_COMB.seq.\*  
66: /cgn2\_6/ptodata/1/pna/US094N\_COMB.seq.\*  
67: /cgn2\_6/ptodata/1/pna/US094O\_COMB.seq.\*  
68: /cgn2\_6/ptodata/1/pna/US094P\_COMB.seq.\*  
69: /cgn2\_6/ptodata/1/pna/US094Q\_COMB.seq.\*  
70: /cgn2\_6/ptodata/1/pna/US094R\_COMB.seq.\*  
71: /cgn2\_6/ptodata/1/pna/US094S\_COMB.seq.\*  
72: /cgn2\_6/ptodata/1/pna/US094T\_COMB.seq.\*  
73: /cgn2\_6/ptodata/1/pna/US094U\_COMB.seq.\*  
74: /cgn2\_6/ptodata/1/pna/US094V\_COMB.seq.\*  
75: /cgn2\_6/ptodata/1/pna/US094W\_COMB.seq.\*  
76: /cgn2\_6/ptodata/1/pna/US094X\_COMB.seq.\*  
77: /cgn2\_6/ptodata/1/pna/US094Y\_COMB.seq.\*  
78: /cgn2\_6/ptodata/1/pna/US094Z\_COMB.seq.\*  
79: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq.\*  
80: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq.\*  
81: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq.\*  
82: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq.\*  
83: /cgn2\_6/ptodata/1/pna/US095E\_COMB.seq.\*  
84: /cgn2\_6/ptodata/1/pna/US095F\_COMB.seq.\*  
85: /cgn2\_6/ptodata/1/pna/US095G\_COMB.seq.\*  
86: /cgn2\_6/ptodata/1/pna/US095H\_COMB.seq.\*  
87: /cgn2\_6/ptodata/1/pna/US095I\_COMB.seq.\*  
88: /cgn2\_6/ptodata/1/pna/US095J\_COMB.seq.\*  
89: /cgn2\_6/ptodata/1/pna/US095K\_COMB.seq.\*  
90: /cgn2\_6/ptodata/1/pna/US095L\_COMB.seq.\*  
91: /cgn2\_6/ptodata/1/pna/US095M\_COMB.seq.\*  
92: /cgn2\_6/ptodata/1/pna/US095N\_COMB.seq.\*  
93: /cgn2\_6/ptodata/1/pna/US095O\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1747	100.0	1747	42	US-09-215-435-66
2	1061	60.7	7983	93	US-60-184-771-693
3	968	55.4	7509	87	US-60-172-360-26327
4	960	55.0	1773	73	US-60-096-116-57
5	909	52.0	1816	54	US-09-471-275-5645
6	908	52.0	2541	34	US-09-009-191-1
7	612	35.0	2370	34	US-09-009-191-3
8	372	21.3	441	54	US-09-465-877-15777
9	331	18.9	447	45	US-09-287-618-34842
10	316	18.1	2838	42	US-09-205-070-13099
11	316	18.1	2838	49	US-09-340-623-13099
12	315	18.0	532	47	US-09-306-609-5482
13	315	18.0	532	92	US-09-522-251-5482

14 314 18.0 449 92 US-09-515-128-1731  
15 268 15.3 477 43 US-09-234-611-1000  
16 260 14.9 311 35 US-09-036-310-4823  
17 258 14.8 417 51 US-09-362-510-19902  
18 237 13.6 237 34 US-09-014-441-814  
19 237 13.6 237 61 US-60-034-841-814  
20 228 13.1 401 44 US-09-274-861-8700  
21 224 12.8 460 54 US-09-465-877-6260  
22 222 12.7 381 92 US-09-489-036-20057  
23 222 12.7 434 45 US-09-280-116-31  
24 218 12.5 278 22 US-08-837-998-1339  
25 218 12.5 278 57 US-60-015-332-1339  
26 212 12.1 313 36 US-09-049-820-5422  
27 212 12.1 313 63 US-60-041-733-1139  
28 204 11.7 246 18 US-08-725-872-974  
29 204 11.7 246 55 US-60-004-952-974  
30 204 11.7 237 35 US-09-023-655-217  
31 197 11.3 248 39 US-09-107-423-2883  
32 197 11.3 248 66 US-60-051-751-2883  
33 191 10.9 198 25 US-08-903-474-505  
34 191 10.9 198 58 US-60-023-339-4469  
35 191 10.9 434 54 US-09-465-877-8393  
36 184 10.5 235 39 US-09-107-425-2189  
37 184 10.5 235 66 US-60-051-751-2189  
38 183 10.5 234 37 US-09-076-667-2272  
39 183 10.5 234 65 US-60-048-002-2272  
40 177 10.1 364 2 US-07-952-911-3919  
41 173 9.9 390 92 US-09-489-036-23801  
42 172 9.8 320 48 US-09-321-214-28221  
43 172 9.8 320 92 US-09-516-333-28221  
44 169 9.7 495 49 US-09-346-956-10403  
45 166 9.5 400 45 US-09-287-618-32169

ALIGNMENTS

RESULT 1  
US-09-215-435-66  
; Sequence 66, Application US/09215435A  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
; APPLICANT: Duclert Aymeric,  
; APPLICANT: Bouqueleret Lydie  
; TITLE OF INVENTION: Extended cDNAs  
; FILE REFERENCE: GENSET.019A  
; CURRENT APPLICATION NUMBER: US/09/215,435A  
; CURRENT FILING DATE: 1998-12-17  
; EARLIER APPLICATION NUMBER: 60/069,957  
; EARLIER FILING DATE: 1997-12-17  
; EARLIER APPLICATION NUMBER: 60/074,121  
; EARLIER FILING DATE: 1998-2-9  
; EARLIER APPLICATION NUMBER: 60/081,563  
; EARLIER FILING DATE: 1998-4-13  
; EARLIER APPLICATION NUMBER: 60/096,116  
; EARLIER FILING DATE: 1998-8-10  
; NUMBER OF SEQ ID NOS: 519  
; SOFTWARE: Patent.pm  
; SEQ ID NO 66  
; LENGTH: 1747  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 10..1062  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 10..57  
; OTHER INFORMATION: Von Heijne matrix  
; OTHER INFORMATION: score 4.9  
; OTHER INFORMATION: seq FYIQAHTLCSG/WS  
; FEATURE:  
; NAME/KEY: polyA\_signal

; LOCATION: 1710..1715  
; FEATURE:  
; NAME/KEY: polyA\_site  
; LOCATION: 1735..1747  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1011  
; OTHER INFORMATION: n=a, g, c or t  
US-09-215-435-66

Query Match 100.0%; Score 1747; DB 42; Length 1747;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gctcaccatgttcccttcattctatctgcaagccactttacactctgtctgggtgg 60  
Db 1 gctcaccatgttcccttcattctatctgcaagccactttacactctgtctgggtgg 60  
QY 61 tcagcacataccgggacctccgggaagggtgtgtatgtgcccacacccagggcaagtgg 120  
Db 61 tcagcacataccgggacctccgggaagggtgtgtatgtgcccacacccagggcaagtgg 120  
QY 121 gaagggagctgggacaccgacctgttaagcatcccccattgcccacacgctcactgtgcgt 180  
Db 121 gaagggagctgggacaccgacctgttaagcatcccccattgcccacacgctcactgtgcgt 180  
QY 181 gccacatctgctgccatcactgaatcagacaagtcttctcaacgggtcccaactgggaa 240  
Db 181 gccacatctgctgccatcactgaatcagacaagtcttctcaacgggtcccaactgggaa 240  
QY 241 ggcatctctgggctgacctatgtgagattgccaggcctgcagagactcccccggagccttcc 300  
Db 241 ggcatctctgggctgacctatgtgagattgccaggcctgcagagactcccccggagccttcc 300  
QY 301 ttgactctctgtgaaagcagaccacgcttcccaacgcttctccctcagactttgtgtgt 360  
Db 301 ttgactctctgtgaaagcagaccacgcttcccaacgcttctccctcagactttgtgtgt 360  
QY 361 gctgggttccccctcaaccagctgaagtgcgtgctctgtcgaggaggagcatgacatt 420  
Db 361 gctgggttccccctcaaccagctgaagtgcgtgctctgtcgaggaggagcatgacatt 420  
QY 421 ggaggtatcgaccactcgtgttacacaggcagctctctgtgtatacaccatccggcgaggag 480  
Db 421 ggaggtatcgaccactcgtgttacacaggcagctctctgtgtatacaccatccggcgaggag 480  
QY 481 tggattatgagtgatgattgtgcgggtggagatcaatggacagagatctgaaaaatggac 540  
Db 481 tggattatgagtgatgattgtgcgggtggagatcaatggacagagatctgaaaaatggac 540  
QY 541 tgcaaggagtacaactatgacaagagcattgtggacagtggaaccacacactctgttg 600  
Db 541 tgcaaggagtacaactatgacaagagcattgtggacagtggaaccacacactctgttg 600  
QY 601 cccaagaaagtgttgaagctgagtcgaatccatcaagggcagctcctccacaggagaag 660  
Db 601 cccaagaaagtgttgaagctgagtcgaatccatcaagggcagctcctccacaggagaag 660  
QY 661 ttccctgacggtttctggctaggagagcagctggtgtgctggcaagcagggcacccct 720  
Db 661 ttccctgacggtttctggctaggagagcagctggtgtgctggcaagcagggcacccct 720  
QY 721 tggaaacattttccagtcactctacactaatgggtgaggttaccacacgactcttc 780  
Db 721 tggaaacattttccagtcactctacactaatgggtgaggttaccacacgactcttc 780  
QY 781 cgcatacaccatctccgcagcaataaccctgcggccagtggaagatgtggccacgtcccaa 840  
Db 781 cgcatacaccatctccgcagcaataaccctgcggccagtggaagatgtggccacgtcccaa 840  
QY 841 gacgactgttacaagtttgccatctcacagtcacacgctcaccggcactgttatggagctgtt 900  
Db 841 gacgactgttacaagtttgccatctcacagtcacacgctcaccggcactgttatggagctgtt 900

```

Db 841 gacgactgttaacaaatttgccatctcaacagtcacagggcactgtttatgggagctgtt 900
Qy 901 atcatggagggtttacgtttgttttcttcttcttcttcttcttcttcttcttcttctt 960
Db 901 atcatggagggtttacgtttgttttcttcttcttcttcttcttcttcttcttcttctt 960
Qy 961 agcgttccatgtgacagatgagttcaggaagcagcagcagcagcagcagcagcagcagc 1020
Db 961 agcgttccatgtgacagatgagttcaggaagcagcagcagcagcagcagcagcagcagc 1020
Qy 1021 ctggacatggaagactgtgggtacaaactccacacagatgagtcacacccctacac 1080
Db 1021 ctggacatggaagactgtgggtacaaactccacacagatgagtcacacccctacac 1080
Qy 1081 catgacctatgtcatgggtgacctgtgacctgtgacctgtgacctgtgacctgtgacctgt 1140
Db 1081 catgacctatgtcatgggtgacctgtgacctgtgacctgtgacctgtgacctgtgacctgt 1140
Qy 1141 gttgacgtggcgtctccctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1200
Db 1141 gttgacgtggcgtctccctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1200
Qy 1201 ctccctgtcgaagtgaagggcccatgggacagatgaggtattccctggaccacact 1260
Db 1201 ctccctgtcgaagtgaagggcccatgggacagatgaggtattccctggaccacact 1260
Qy 1261 ccgtgttccactttggttcacaaagttaggagacagatggcactgtggccagagcacctc 1320
Db 1261 ccgtgttccactttggttcacaaagttaggagacagatggcactgtggccagagcacctc 1320
Qy 1321 aggaacccctccacacacacacacacacacacacacacacacacacacacacacacacac 1380
Db 1321 aggaacccctccacacacacacacacacacacacacacacacacacacacacacacacac 1380
Qy 1381 tgggttccagggactgtacctgtaggagacagaaagaaagaaagaaagaaagaaagaaag 1440
Db 1381 tgggttccagggactgtacctgtaggagacagaaagaaagaaagaaagaaagaaagaaag 1440
Qy 1441 cgggaatactctgttgcacctcaaatatttaagtcgggaaattctgtctgaaacttcag 1500
Db 1441 cgggaatactctgttgcacctcaaatatttaagtcgggaaattctgtctgaaacttcag 1500
Qy 1501 cctgaaccccttgcacacacacacacacacacacacacacacacacacacacacacacac 1560
Db 1501 cctgaaccccttgcacacacacacacacacacacacacacacacacacacacacacacac 1560
Qy 1561 gtttcagaagtagtgcacacacacacacacacacacacacacacacacacacacacacac 1620
Db 1561 gtttcagaagtagtgcacacacacacacacacacacacacacacacacacacacacacac 1620
Qy 1621 gcagagaagagaccagctgtttccctgctgctgctgctgctgctgctgctgctgctgctg 1680
Db 1621 gcagagaagagaccagctgtttccctgctgctgctgctgctgctgctgctgctgctgctg 1680
Qy 1681 ttgtctattgttttagagacagggactgtataaacaacacacacacacacacacacacac 1740
Db 1681 ttgtctattgttttagagacagggactgtataaacaacacacacacacacacacacacac 1740
Qy 1741 aaaaaa 1747
Db 1741 aaaaaa 1747

```

RESULT 2

US-60-184-771-693  
; Sequence 693, Application US/60184771  
; GENERAL INFORMATION:

; APPLICANT: Hodgson, David M.  
; APPLICANT: Lincoln, Stephen E.  
; APPLICANT: Russo, Frank D.  
; APPLICANT: Spiro, Peter A.  
; APPLICANT: Banville, Steve C.  
; APPLICANT: Bratcher, Shawn R.  
  
; Jones, Anissa L.  
; Yu, Jimmy Y.  
; Greenawalt, Lila B.  
; Panzer, Scott R.  
; Roseberry, Ann M.  
; Wright, Rachel J.

```

; APPLICANT: Dufour, Gerard E.
; APPLICANT: Cohen, Howard J.
; APPLICANT: Rosen, Bruce
; APPLICANT: Shah, Purvi
; APPLICANT: Chalup, Michael S.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PT-0120 P
; CURRENT APPLICATION NUMBER: US/60/184.771
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 1244
; SOFTWARE: PERL Program
; SEQ ID NO 693
; LENGTH: 7983
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 234216.15.1
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 3518
; OTHER INFORMATION: a, t, c, g, or other
US-60-184-771-693

```

```

Query Match 60.7%; Score 1061; DB 93; Length 7983;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1671; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

Qy 60 gtccagacacatccgggacccctccgaaagggtgtgtatgtgcctacacccagggaagtgt 119
Db 248 gtccagacacatccgggacccctccgaaagggtgtgtatgtgcctacacccagggaagtgt 307
Qy 120 ggaaggaggagctgggacccgacctgttaagcatcccccacatgcccacacacacacacac 179
Db 308 ggaaggaggagctgggacccgacctgttaagcatcccccacatgcccacacacacacacac 367
Qy 180 tgcacacattgtgcctacatgaatcagacaaagtcttctcatcaacggtcccaactggga 239
Db 368 tgcacacattgtgcctacatgaatcagacaaagtcttctcatcaacggtcccaactggga 427
Qy 240 aggcactcctgggctggcctatgtcagatgcagagcctgacgactcccccagacacatt 299
Db 428 aggcactcctgggctggcctatgtcagatgcagagcctgacgactcccccagacacatt 487
Qy 300 ctttgactctctgttaaagcagacccacgttcccaacctctctccctgcagctttgtgg 359
Db 488 ctttgactctctgttaaagcagacccacgttcccaacctctctccctgcagctttgtgg 547
Qy 360 tgcctggcttcccccctcaaccagctcgaagtgtgcctctgtcggaggagacatgatcat 419
Db 548 tgcctggcttcccccctcaaccagctcgaagtgtgcctctgtcggaggagacatgatcat 607
Qy 420 tggaggtatgcacacactgcgtgtacacaggaagctcgtgtgtgtgtgtgtgtgtgtgtgt 479
Db 608 tggaggtatgcacacactgcgtgtacacaggaagctcgtgtgtgtgtgtgtgtgtgtgtgt 667
Qy 480 gtgggtattatgaggtgatcattgtgcgggtggagatcaatggacagagatctgaaatgga 539
Db 668 gtgggtattatgaggtgatcattgtgcgggtggagatcaatggacagagatctgaaatgga 727
Qy 540 ctgcaaggaggtacaaactatgacaaagacattgtgtgacagtggcaccacacacacacacac 599
Db 728 ctgcaaggaggtacaaactatgacaaagacattgtgtgacagtggcaccacacacacacacac 787
Qy 600 gcccaaaaaagtgtttgaagctgcagtcacaaatccatcaaggcagcctcctccacaggaaa 659
Db 788 gcccaaaaaagtgtttgaagctgcagtcacaaatccatcaaggcagcctcctccacaggaaa 847
Qy 660 gtccctcactcgtttctgtgtaggagacagcagctgtgtgtgtgtgtgtgtgtgtgtgtgt 719
Db 848 gtccctcactcgtttctgtgtaggagacagcagctgtgtgtgtgtgtgtgtgtgtgtgtgt 907

```

```

; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; FILE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polym
; FILE REFERENCE: GX-0007 P
; CURRENT APPLICATION NUMBER: US/60/172,360
; CURRENT FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 29838
; SOFTWARE: PERL Program
; SEQ ID NO 26327
; LENGTH: 7509
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: incyte ID No: 234216.12
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 100
; OTHER INFORMATION: a, t, c, g, or other
; 15-60-172-360-26327

```

Query Match	Score 968;	DB 87;	Length 7509;
Best Local Similarity	55.4%;		
Matches 1358. Conservative	99.6%;		
Pred. No. 0;			
0: Mismatches			
3: Indels			
2: Gaps			

Query Match	55.4%;	Score 968;	DB 87;	Length 7509;
Best Local Similarity	99.6%;	Pred. No. 0;		Gaps
Matches 1358;	Conservative	Mismatches 0;		
QY	375	caaccagctctgaagtgctggccctctctcgaggaggagcatgacattgagggtatcgacca	434	
Db	101	caacagctctgaagtgctggccctctctcgaggaggagcatgacattgagggtatcgacca	160	
QY	435	ctcgcgtgtacacaggcagctctctgtgtatacacaccatccgcgggagtggtattatgaggt	494	
Db	161	ctcgcgtgtacacaggcagctctctgtgtatacacaccatccgcgggagtggtattatgaggt	220	
QY	495	gatcattgtcggtgtgagatcaatggacaggatctgaaatggactcgcaaggagtacaa	554	
Db	221	gatacttgtcggtgtgagatcaatggacaggatctgaaatggactcgcaaggagtacaa	280	
QY	555	ctatgacaagagcattgtgacagtggcacacacacctctgttcgcccaagaagtgtt	614	
Db	281	ctatgacaagagcattgtgacagtggcacacacacctctgttcgcccaagaagtgtt	340	
QY	615	tgaactcagctcaaatccatcaagagcagctctctccacggagaaagtctccctgacggttt	674	
Db	341	tgaactcagctcaaatccatcaagagcagctctctccacggagaaagtctccctgacggttt	400	
QY	675	ctgggtagagagcagctggtgtgtcggaagcaggcaccaccccttggaacattttccc	734	
Db	401	ctgggtagagagcagctggtgtgtcggaagcaggcaccaccccttggaacattttccc	460	
QY	735	agtcattcactctactctaagtgtgaggtattaccaaccagctcttcgcatacaccatctt	794	
Db	461	agtcattcactctactctaagtgtgaggtattaccaaccagctcttcgcatacaccatctt	520	
QY	795	tcgcgagcaataccttcggcgcagtggaagtgtggccacgtcccaagaagcactgtttacaa	854	
Db	521	tcgcgagcaataccttcggcgcagtggaagtgtggccacgtcccaagaagcactgtttacaa	580	
QY	855	gtttggcatctcacagtcataccaaggcactgttatgggagctgttatactgaggggctt	914	
Db	581	gtttggcatctcacagtcataccaaggcactgttatgggagctgttatactgaggggctt	640	
QY	915	ctacgttgtcttttgatcgggcccgaacaaatggctttgtgtctgcagcgttgccatgtt	974	
Db	641	ctacgttgtcttttgatcgggcccgaacaaatggctttgtgtctgcagcgttgccatgtt	700	
QY	975	gcacatgatgattcaggacggcagcgttggaaggccctttgtcactctggacatggaag	103	
Db	701	gcacatgatgattcaggacggcagcgttggaaggccctt-tttgtcactctggacatggaag	759	

QY	720	ttggaacattttccagtcac	tctcactctac	cttaacta	atgggtgaggttacc	accagtcctt	779	
Db	908	ttggaacattttccagtcac	tctcactctac	cttaacta	atgggtgaggttacc	accagtcctt	967	
QY	780	ccgcatcaccatcttcgcg	agcaatac	ctctgcgcagtcg	gcgcagtcggaagat	gtgccacgtccca	839	
Db	968	ccgcatcaccatcttcgcg	agcaatac	ctctgcgcagtcg	gcgcagtcggaagat	gtgccacgtccca	1027	
QY	840	agacgactgttacaa	gtttgcacat	ctcactctacagtc	ctcacggtgcactgt	tatagggagctgt	899	
Db	1028	agacgactgttacaa	gtttgcacat	ctcactctacagtc	ctcacggtgcactgt	tatagggagctgt	1087	
QY	900	latcatgagggcttctac	gttctgtctt	tgatc-gggcccgaa	aaacgaat	tggtgcttgc	958	
Db	1088	latcatgagggcttctac	gttctgtctt	tgatc-gggcccgaa	aaacgaat	tggtgcttgc	1147	
QY	959	tcagcgcttgccatgt	gcagatgatt	tcacagacgagcg	gtggaagccn	ttttgtc	1018	
Db	1148	tcagcgcttgccatgt	gcagatgatt	tcacagacgagcg	gtggaagccn	ttttgtc	1206	
QY	1019	accttggacatggaag	actgtgtgc	atacaacatt	tcacagacagat	tgatcgaacccctcatg	1078	
Db	1207	accttggacatggaag	actgtgtgc	atacaacatt	tcacagacagat	tgatcgaacccctcatg	1266	
QY	1079	acctatgactatgtcat	ggctgccc	ctctgcgcctcttc	atgctgcactct	tgccctcatg	1138	
Db	1267	acctatgactatgtcat	ggctgccc	ctctgcgcctcttc	atgctgcactct	tgccctcatg	1326	
QY	1139	gtgtgcagtggcgct	gcctgcgctgcgc	gcagacagat	tgatgacttt	gtctgatgac	1198	
Db	1327	gtgtgcagtggcgct	gcctgcgctgcgc	gcagacagat	tgatgacttt	gtctgatgac	1386	
QY	1199	atctccctctgaagt	gagagggcccat	gggcagagat	taggattcccc	ctggaccacac	1258	
Db	1387	atctccctctgaagt	gagagggcccat	gggcagagat	taggattcccc	ctggaccacac	1446	
QY	1259	ctccgtgtgttcaact	ttgttcacaa	gtagtagacagat	tgggcacct	tggtgccagacacc	1318	
Db	1447	ctccgtgtgttcaact	ttgttcacaa	gtagtagacagat	tgggcacct	tggtgccagacacc	1506	
QY	1319	tcaggacctcccccac	cccccaaat	ccctctgccttgat	ggagaggaagaa	ggctgggca	1378	
Db	1507	tcaggacctcccccac	cccccaaat	ccctctgccttgat	ggagaggaagaa	ggctgggca	1566	
QY	1379	ggtgggttcagg	ggactgtacct	gttaggagacagaaa	gagaaagac	actctgct	1438	
Db	1567	ggtgggttcagg	ggactgtacct	gttaggagacagaaa	gagaaagac	actctgct	1626	
QY	1439	ggcggaatactctt	ggtaacctcaaa	tttaagt	cgggaaattct	ctgctgaaacttc	1498	
Db	1627	ggcggaatactctt	ggtaacctcaaa	tttaagt	cgggaaattct	ctgctgaaacttc	1686	
QY	1499	agccctgaacctttgt	-cacca	ttccctttaaatt	ctccacccaa	agattctcttttc	1557	
Db	1687	agccctgaacctttgt	-cacca	ttccctttaaatt	ctccacccaa	agattctcttttc	1746	
QY	1558	ttagtttcagaag	tacttggc	atacacgcaggt	tacacttggc	gtgtccctgtgtgac	1617	
Db	1747	ttagtttcagaag	tacttggc	atacacgcaggt	tacacttggc	gtgtccctgtgtgac	1806	
QY	1618	ctggcagagaag	agaccaa	gctgtttccct	gctggccaa	agtcagtaggagagatgca	1677	
Db	1807	ctggcagagaag	agaccaa	gctgtttccct	gctggccaa	agtcagtaggagagatgca	1866	
QY	1678	cagtttgcta	ttgtcttt	tagagacagggact	gtataa	caagccta	cattggtgc	1736
Db	1867	cagtttgcta	ttgtcttt	tagagacagggact	gtataa	caagccta	cattggtgc	1925
RESULT 3								
6 172-360-2627								

```

QY 1035 actgtggtacaaattccacacagacagatgagtcgaacccctcatgacacatagcctatgtca 1094
Db 760 actgtggtacaaattccacacagacagatgagtcgaacccctcatgacacatagcctatgtca 819
QY 1095 tggctgcccattcgccctttcatctgccaactctgcccctcatggtgtgcagtgccgct 1154
Db 820 tggctgcccattcgccctttcatctgccaactctgcccctcatggtgtgcagtgccgct 879
QY 1155 gctccgctgctgcccagcagcatgatgactttgtgatgacatccctcctgctgaagt 1214
Db 880 gctccgctgctgcccagcagcatgatgactttgtgatgacatccctcctgctgaagt 939
QY 1215 gagagggccatgggcagaaatagagattccctcctggaccacacccctcctggttcacttt 1274
Db 940 gagagggccatgggcagaaatagagattccctcctggaccacacccctcctggttcacttt 999
QY 1275 ggtcaccaagttagagacacagatggcacctgtggccagagacacctcaggaccctccccc 1334
Db 1000 ggtcaccaagttagagacacagatggcacctgtggccagagacacctcaggaccctccccc 1059
QY 1335 ccaccaaatgctctgcttgccttgatgggaagaaaggctggcaagtggtggttccagggac 1394
Db 1060 ccaccaaatgctctgcttgccttgatgggaagaaaggctggcaagtggtggttccagggac 1119
QY 1395 tgtacctgtaggagacagaaaagagaagaaagacactctgctgcccgggaataactcttg 1454
Db 1120 tgtacctgtaggagacagaaaagagaagaaagacactctgctgcccgggaataactcttg 1179
QY 1455 gtacacctcaaatattagtcgggaattctgctgcttgaacttcagccctcagaccccttgt 1514
Db 1180 gtacacctcaaatattagtcgggaattctgctgcttgaacttcagccctcagaccccttgt 1239
QY 1515 -caccattccttaaatctccacccaaagtattcttctttagtttcagaagtac 1573
Db 1240 ccaccattccttaaatctccacccaaagtattcttctttagtttcagaagtac 1299
QY 1574 tggcatcacacagaggttaacttggcgtgtgtccctgtgtacccctggcagagaagagac 1633
Db 1300 tggcatcacacagaggttaacttggcgtgtgtccctgtgtacccctggcagagaagagac 1359
QY 1634 caagctgtttcctgctggtccaaagtcagtaggagagatgcacagtttgcatttgc 1693
Db 1360 caagctgtttcctgctggtccaaagtcagtaggagagatgcacagtttgcatttgc 1419
QY 1694 ttagagacagggactgtataacaaagcctaacttggtgcaaa 1736
Db 1420 ttagagacagggactgtataacaaagcctaacttggtgcaaa 1462

```

RESULT 4

```

US-60-096-116-57
; Sequence 57, Application US/60096116
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste Dumas Milne
; APPLICANT: Duclert, Aymeric
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SECRETED PROTEINS
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESS: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/096,116
; FILING DATE:
; CLASSIFICATION:

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.038PR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1773 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Substantia nigra
; NAME/KEY: sig_peptide
; LOCATION: 10..57
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 4.9
; OTHER INFORMATION: seq FLYQAHTLCGG/WS
; FEATURE:
; NAME/KEY: poly_a_signal
; LOCATION: 1712..1717
; FEATURE:
; NAME/KEY: poly_a
; LOCATION: 1736..1749
; IDENTIFICATION METHOD: blastn2
; FEATURE:
; NAME/KEY: est
; LOCATION: 1566..1735
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 100
; OTHER INFORMATION: region 1..170
; OTHER INFORMATION: id R12815
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(1569..1735)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 100
; OTHER INFORMATION: region 11..177
; OTHER INFORMATION: id AA65039
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(1570..1735)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 100
; OTHER INFORMATION: region 24..189
; OTHER INFORMATION: id AA758721
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(1292..1735)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 99
; OTHER INFORMATION: region 20..464
; OTHER INFORMATION: id AA701598
; FEATURE:
; NAME/KEY: est
; LOCATION: complement(1328..1735)
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 99
; OTHER INFORMATION: region 20..427
; OTHER INFORMATION: id AA677847
; FEATURE:
; NAME/KEY: est
; LOCATION: 1391..1699
; IDENTIFICATION METHOD: blastn2
; OTHER INFORMATION: identity 99
; OTHER INFORMATION: region 1..310
; OTHER INFORMATION: id HSC3HD121
; FEATURE:

```

```

NAME/KEY: est
LOCATION: complement(1429..1735)
IDENTIFICATION METHOD: blastn2
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 20..327
OTHER INFORMATION: 1d AA705405
FEATURE:

NAME/KEY: est
LOCATION: complement(1436..1735)
IDENTIFICATION METHOD: blastn2
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 1..301
OTHER INFORMATION: 1d AA729830
FEATURE:

NAME/KEY: est
LOCATION: 1469..1735
IDENTIFICATION METHOD: blastn2
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 1..267
OTHER INFORMATION: 1d T75407
FEATURE:

NAME/KEY: est
LOCATION: complement(1473..1735)
IDENTIFICATION METHOD: blastn2
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 39..302
OTHER INFORMATION: 1d AA970909
FEATURE:

NAME/KEY: est
LOCATION: 1391..1649
IDENTIFICATION METHOD: blastn2
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 1..260
OTHER INFORMATION: 1d HSC02A051
FEATURE:

NAME/KEY: est
LOCATION: 1481..1735
IDENTIFICATION METHOD: blastn2
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 1..256
OTHER INFORMATION: 1d T08924
FEATURE:

NAME/KEY: est
LOCATION: complement(1524..1735)
IDENTIFICATION METHOD: blastn2
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 6..217
OTHER INFORMATION: 1d AA063085
FEATURE:

NAME/KEY: est
LOCATION: complement(1235..1505)
IDENTIFICATION METHOD: blastn2
OTHER INFORMATION: identity 98
OTHER INFORMATION: region 1..271
OTHER INFORMATION: 1d AA918558
FEATURE:

NAME/KEY: est
LOCATION: 1279..1469
IDENTIFICATION METHOD: blastn2
OTHER INFORMATION: identity 98
OTHER INFORMATION: region 7..197
OTHER INFORMATION: 1d AA063135
FEATURE:

NAME/KEY: est
LOCATION: 1566..1735
IDENTIFICATION METHOD: blastn2
OTHER INFORMATION: identity 98
OTHER INFORMATION: region 1..170
OTHER INFORMATION: 1d T89769
FEATURE:

NAME/KEY: est
LOCATION: complement(1311..1735)
IDENTIFICATION METHOD: blastn2

```

[illegible]





QY 244 atcctggggctggcctatgctgagattgccaaggcctgaagactcccgagcctttcttt 303  
 Db 71 atcctggggctggcctatgctgagattgccaaggcctgaagactcccgagcctttcttt 130  
 QY 304 gactctctggttaagcagacacacagcttcccaacctcttctccctgcagctttggtgct 363  
 Db 131 gactctctggttaagcagacacacagcttcccaacctcttctccctgcagctttggtgct 190  
 QY 364 ggtctcccccctcaacacagctggaagtgtgctgctctgtcgagggagcagatgacatgga 423  
 Db 191 ggtctcccccctcaacacagctggaagtgtgctgctctgtcgagggagcagatgacatgga 250  
 QY 424 ggtatcgaccactcgtgtacacagcagctggaagtgtgctgctctgtcgagggagcagatgga 483  
 Db 251 ggtatcgaccactcgtgtacacagcagctggaagtgtgctgctctgtcgagggagcagatgga 310  
 QY 484 tattatgaggtgacatgtgtggtggtgagatcaatggacaggaatctgaaatggactgc 543  
 Db 311 tattatgaggtgacatgtgtggtggtgagatcaatggacaggaatctgaaatggactgc 370  
 QY 544 aaggagtacaactatgacaagagcattgtgacagtggtgacacacacacacactcgtttgccc 603  
 Db 371 aaggagtacaactatgacaagagcattgtgacagtggtgacacacacacacactcgtttgccc 430  
 QY 604 aagaaagtgttgaagctgacgtcaaatccatcaaggcagcctcctccacggagagttc 663  
 Db 431 aagaaagtgttgaagctgacgtcaaatccatcaaggcagcctcctccacggagagttc 490  
 QY 664 cctgacggtttctggatgagagcagctggtgtgctggcagcaggcgaacaccccttgg 723  
 Db 491 cctgaggtttctggatgagagcagctggtgtgctggcagcaggcgaacaccccttgg 550  
 QY 724 aacatttccagctacatcctcctacctaattgggtgaggttaccacacacacacacacac 783  
 Db 551 aacatttccagctacatcctcctacctaattgggtgaggttaccacacacacacacacac 610  
 QY 784 atcacatccttcgcagcaatacctcgcgcagtcagtggaagatgtggccacgctcccaaac 843  
 Db 611 atcacatccttcgcagcaatacctcgcgcagtcagtggaagatgtggccacgctcccaaac 670  
 QY 844 gactgttaagtttgcctatcctacagtcacacgagcagctggttatggagctgtatc 903  
 Db 671 gactgttaagtttgcctatcctacagtcacacgagcagctggttatggagctgtatc 730  
 QY 904 atgagggcttctacgttcttctgtatcgccgcccgaagaaatggcttctgtctcagc 963  
 Db 731 atgagggcttctacgttcttctgtatcgccgcccgaagaaatggcttctgtctcagc 790  
 QY 964 gcttgccatgtcacagatgagttcagcagcagcagcagcagcagcagcagcagcagcagc 1023  
 Db 791 gcttgccatgtcacagatgagttcagcagcagcagcagcagcagcagcagcagcagcagc 849  
 QY 1024 ggcacatggaagactggtgctacacattccacacagacagatgagtcacacacacacacac 1083  
 Db 850 ggcacatggaagactggtgctacacattccacacagacagatgagtcacacacacacacac 909  
 QY 1084 agcctatgtcatggtgctgctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1143  
 Db 910 agcctatgtcatggtgctgctcctcctcctcctcctcctcctcctcctcctcctcctcct 969  
 QY 1144 tcagtggtgctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1203  
 Db 970 tcagtggtgctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1029  
 QY 1204 cctgctgaagtgaagagcccatgggcagagatgaggttccctcctggacacacacacac 1263  
 Db 1030 cctgctgaagtgaagagcccatgggcagagatgaggttccctcctggacacacacacac 1089  
 QY 1264 tgggttcacttgggtcacaagttaggagacacagatggcagcctgtggccagagcagcctcag 1323  
 Db 1090 tgggttcacttgggtcacaagttaggagacacagatggcagcctgtggccagagcagcctcag 1149

QY 1324 accctcccccaccccaaatgctctgctgctgaggaagaaagagcctggcagagtggtg 1383  
 Db 1150 accctcccccaccccaaatgctctgctgctgaggaagaaagagcctggcagagtggtg 1209  
 QY 1384 gttccagggactgtaccctgtagagacagaaagaaagaaagacactctgctggtggtg 1443  
 Db 1210 gttccagggactgtaccctgtagaagaaagaaagaaagacactctgctggtggtg 1269  
 QY 1444 gaatactctggtcacctcaaatttaagtcgggaaatctctgctgctgaaacttcagccc 1503  
 Db 1270 gaatactctggtcacctcaaatttaagtcgggaaatctctgctgctgaaacttcagccc 1329  
 QY 1504 tgaactttgt-cacactcctttaaattcctccaccccaaaagtattctcttttttagt 1562  
 Db 1330 tgaactttgtccaccattcctttaaattcctccaccccaaaagtattctcttttttagt 1389  
 QY 1563 ttcaagaactgtgacacacacagcaggttaccttggtgctggtggtggtggtggtggtg 1622  
 Db 1390 ttcaagaactgtgacacacacagcaggttaccttggtgctggtggtggtggtggtggtg 1449  
 QY 1623 agagaagagaccagctgtttccctgctggtggtggtggtggtggtggtggtggtggtg 1682  
 Db 1450 agagaagagaccagctgtttccctgctggtggtggtggtggtggtggtggtggtggtg 1509  
 QY 1683 tgctattt 1690  
 Db 1510 tgctattt 1517

RESULT 6  
 US-09-009-191-1  
 ; Sequence 1, Application US/09009191  
 ; GENERAL INFORMATION:  
 ; APPLICANT: POWELL, DAVID  
 ; APPLICANT: CHAPMAN, CONRAD  
 ; APPLICANT: MURPHY, KAY  
 ; APPLICANT: SMITH, TRUDI  
 ; TITLE OF INVENTION: ASP2  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: RATNER & PRESTIA  
 ; STREET: P.O. BOX 980  
 ; CITY: VALLEY FORGE  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19482  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/009,191  
 ; FILING DATE: 20-JAN-1998  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: UK 9701684.4  
 ; FILING DATE: 28-JAN-1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: PRESTIA, PAUL F  
 ; REGISTRATION NUMBER: 23,031  
 ; REFERENCE/POCKET NUMBER: GH-70368  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 610-407-0700  
 ; TELEFAX: 610-407-0701  
 ; TELEX: 846169  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2541 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA





```
Db 15 acggcccaactgggaagcaccctggggtggtcctgctgagattgcagcctgacg 74
QY 284 actcccgagcctctcttctgactctctgtaagcagaccacacgctcccaacctctct 343
Db 75 actccctgagcctctcttctgactctctgtaagcagaccacacgctcccaacctctct 134
QY 344 cctgcagcttctggtgctgcttccctccctcaaccagctctgaagctgctgctctgctg 403
Db 135 cctgcagcttctggtgctgcttccctccctcaaccagctctgaagctgctgctctgctg 194
QY 404 gaggagcatgatcattgagatgacacactcgtgtacacaggcagctctggtata 463
Db 195 gggggagcatgacattgaggtatcgacacactcgtgtacacaggcagctctggtata 254
QY 464 caccatccggcgggagtggtattatgagtgatcattgtcggtggagatcaatggac 523
Db 255 caccatccggcgggagtggtattatgaggtcatcattgtcggtggagatcaatggac 314
QY 524 aggatcgaatgagctgcaaggaggtacaaactatgacaagagcattgtggacagtgga 583
Db 315 aggatcgaatgagctgcaaggaggtacaaactatgacaagagcattgtggacagtgga 374
QY 584 ccaccaaccttggttggccagaaagtggttgaagctcagtcacaaatccatcaaggcag 643
Db 375 ccaccaaccttggttggccagaaagtggttgaagctcagtcacaaatccatcaaggcag 434
QY 644 cctctccacacgga 656
Db 435 cctctccacacgga 447
```

## RESULT 10

US-09-205-070-13099

; Sequence 13099, Application US/09205070

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA

; FILE OF INVENTION: LIBRARIES

; FILE REFERENCE: 20411-748

; CURRENT APPLICATION NUMBER: US/09/205,070

; EARLIER FILING DATE: 1998-12-03

; NUMBER OF SEQ ID NOS: 45207

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 13099

; LENGTH: 2838

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(2838)

; OTHER INFORMATION: n = A,T,C or G

US-09-205-070-13099

```
Query Match 18.1%; Score 316; DB 42; Length 2838;
Best Local Similarity 99.6%; Pred. No. 1.6e-139;
Matches 486; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
```

```
QY 1250 ggaccacacctcgtggttcacttggtcacaagttaggagacacagatggcaccctgtgc 1309
Db 467 ggaccacacctcgtggttcacttggtcacaagttaggagacacagatggcaccctgtgc 526
QY 1310 cagagcacctcaggaccctcccccaccacaaatgccttgccttgatgagaaagaaa 1369
Db 527 cagagcacctcaggaccctcccccaccacaaatgccttgccttgatgagaaagaaa 586
QY 1370 ggcctggaaggtgggttccaggagctgtacctgtagggacagaaaaagagaagaag 1429
Db 587 ggcctggaaggtgggttccaggagctgtacctgtagggacagaaaaagagaagaag 646
QY 1430 cactctgctgggggaataactcttggtcaccccaatttaagtgcgggaattctgctgt 1489
```

```
Db 647 cactctgctggcggaataactcttggtcacccctcaaatttaagtcgggaaattctctgct 706
QY 1490 tgaacttcagccctgaacctttgt-cacattctctttaaattctccaacccaagtatt 1548
Db 707 tgaacttcagccctgaacctttgtccacattctctttaaattctccaacccaagtatt 766
QY 1549 ctctcttctttagtttcagaagtactgcatcacacgaggttaacttggcgtgtgtccc 1608
Db 767 ctctcttctttagtttcagaagtactgcatcacacgaggttaacttggcgtgtgtccc 826
QY 1609 tctgtacccttcagagagaagacaagcttcttccctgctgcccgaagctcagtagca 1668
Db 827 tctgtacccttcagagagaagacaagcttcttccctgctgcccgaagctcagtagca 886
QY 1669 gaggtgacacagtttctgatttctttagagacagggagctgtataaacaagcctaact 1728
Db 887 gaggtgacacagtttctgatttctttagagacagggagctgtataaacaagcctaact 946
QY 1729 ggtgcaaa 1736
Db 947 ggtgcaaa 954
```

## RESULT 11

US-09-340-623-13099

; Sequence 13099, Application US/09340623

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA

; FILE OF INVENTION: LIBRARIES

; FILE REFERENCE: 20411-748CON1

; CURRENT APPLICATION NUMBER: US/09/340,623

; EARLIER FILING DATE: 1999-06-28

; NUMBER OF SEQ ID NOS: 45207

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 13099

; LENGTH: 2838

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(2838)

; OTHER INFORMATION: n = A,T,C or G

US-09-340-623-13099

```
Query Match 18.1%; Score 316; DB 49; Length 2838;
Best Local Similarity 99.6%; Pred. No. 1.6e-139;
Matches 486; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
```

```
QY 1250 ggaccacacctcgtggttcacttggtcacaagttaggagacacagatggcaccctgtgc 1309
Db 467 ggaccacacctcgtggttcacttggtcacaagttaggagacacagatggcaccctgtgc 526
QY 1310 cagagcacctcaggaccctcccccaccacaaatgccttgccttgatgagaaagaaa 1369
Db 527 cagagcacctcaggaccctcccccaccacaaatgccttgccttgatgagaaagaaa 586
QY 1370 ggcctggaaggtgggttccaggagctgtacctgtagggacagaaaaagagaagaag 1429
Db 587 ggcctggaaggtgggttccaggagctgtacctgtagggacagaaaaagagaagaag 646
QY 1430 cactctgctgggggaataactcttggtcaccccaatttaagtgcgggaattctgctgt 1489
Db 647 cactctgctgggggaataactcttggtcaccccaatttaagtgcgggaattctgctgt 706
QY 1490 tgaacttcagccctgaacctttgt-cacattctctttaaattctccaacccaagtatt 1548
Db 707 tgaacttcagccctgaacctttgtccaccattctctttaaattctccaacccaagtatt 766
QY 1549 ctctcttctttagtttcagaagtactggtcagacacgcaggttaacttggcgtgtgtccc 1608
```

Db 767 cttcttttttagtttcagaaagtaactggtcaccacagcaggtttacottggtgtgtccc 826  
 Qy 1609 tgggtaccctggcagagaagacacaaagctgtttccctgctggccaaagtacagtagga 1668  
 Db 827 tgggtaccctggcagagaagacacaaagctgtttccctgctggccaaagtacagtagga 886  
 Qy 1669 gaggatcacagtttctattgttttagacagaggactgtataaacaagcctaaccatt 1728  
 Db 887 gaggatcacagtttctattgttttagacagaggactgtataaacaagcctaaccatt 946  
 Qy 1729 ggtgcaaa 1736  
 Db 947 ggtgcaaa 954

RESULT 12  
 US-09-306-609-5482  
 ; Sequence 5482, Application US/09306609  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hyseq, Inc.  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
 ; FILE REFERENCE: 20411-718CON2  
 ; CURRENT APPLICATION NUMBER: US/09/306,609  
 ; EARLIER FILING DATE: 1999-05-07  
 ; EARLIER APPLICATION NUMBER: US 09/168,296  
 ; EARLIER FILING DATE: 1998-10-07  
 ; EARLIER APPLICATION NUMBER: US 08/949,553  
 ; EARLIER FILING DATE: 1997-10-09  
 ; NUMBER OF SEQ ID NOS: 13025  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 5482  
 ; LENGTH: 532  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(532)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-306-609-5482

Query Match 18.0%; Score 315; DB 47; Length 532;  
 Best Local Similarity 99.8%; Pred. No. 4.2e-139;  
 Matches 435; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 Qy 788 ccactctccgcagcaatacctcgccagtggaagatggtgccacgtgcccaagacgact 847  
 Db 1 ccactctccgcagcaatacctcgccagtggaagatggtgccacgtgcccaagacgact 60  
 Qy 848 gttacaagtttgccatctcacagtcacacgggcagctgttatggagctgttatcatgg 907  
 Db 61 gttacaagtttgccatctcacagtcacacgggcagctgttatggagctgttatcatgg 120  
 Qy 908 aggggttttaagttgtcttttagtcggcccgaaacgaattgcttgcctgaagcgtt 967  
 Db 121 aggggttttaagttgtcttttagtcggcccgaaacgaattgcttgcctgaagcgtt 180  
 Qy 968 gccatgtgcacgatgagttcaggacggcagcgggtggaagcccttttgcaccttgac 1027  
 Db 181 gccatgtgcacgatgagttcaggacggcagcgggtggaagcccttttgcaccttgac 239  
 Qy 1028 atgggaagacttggttacaacattccacagacagatgagtaaacctcatgaccatagcc 1087  
 Db 240 atgggaagacttggttacaacattccacagacagatgagtaaacctcatgaccatagcc 299  
 Qy 1088 tatgtcatggtgcacatcgccctctttcatgtgcacctgtgctcatggtgtgtcag 1147  
 Db 300 tatgtcatggtgcacatcgccctctttcatgtgcacctgtgctcatggtgtgtcag 359  
 Qy 1148 tggcgctgcctcgcctgctgcgcagcagcatgatgtatttgcctgtgatgacatccctcg 1207

Db 360 tggcgctgcctcgcctgctgcgcagcagcatgatgactttgctgtatgacatccctcg 419  
 Qy 1208 ctgaagttagggagcc 1223  
 Db 420 ctgaagttagggagcc 435  
 RESULT 13  
 US-09-522-251-5482  
 ; Sequence 5482, Application US/09522251  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Drmanac, Radoje T.  
 ; APPLICANT: Stache-Grain, Birgit  
 ; APPLICANT: Dickson, Mark C.  
 ; APPLICANT: Jones, Lee W.  
 ; APPLICANT: Garcia, Veronica E.  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
 ; FILE REFERENCE: 718CIP  
 ; CURRENT APPLICATION NUMBER: US/09/522,251  
 ; CURRENT FILING DATE: 2000-03-09  
 ; EARLIER APPLICATION NUMBER: 09/306,609  
 ; EARLIER FILING DATE: 1999-05-07  
 ; EARLIER APPLICATION NUMBER: 09/168,296  
 ; EARLIER FILING DATE: 1998-10-07  
 ; EARLIER APPLICATION NUMBER: 60/150,686  
 ; EARLIER FILING DATE: 1997-10-09  
 ; NUMBER OF SEQ ID NOS: 13025  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 5482  
 ; LENGTH: 532  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(532)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-522-251-5482

Query Match 18.0%; Score 315; DB 92; Length 532;  
 Best Local Similarity 99.8%; Pred. No. 4.2e-139;  
 Matches 435; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 Qy 788 ccactctccgcagcaatacctcgccagtggaagatggtgccacgtgcccaagacgact 847  
 Db 1 ccactctccgcagcaatacctcgccagtggaagatggtgccacgtgcccaagacgact 60  
 Qy 848 gttacaagtttgccatctcacagtcacacgggcagctgttatggagctgttatcatgg 907  
 Db 61 gttacaagtttgccatctcacagtcacacgggcagctgttatggagctgttatcatgg 120  
 Qy 908 aggggttttaagttgtcttttagtcggcccgaaacgaattgcttgcctgaagcgtt 967  
 Db 121 aggggttttaagttgtcttttagtcggcccgaaacgaattgcttgcctgaagcgtt 180  
 Qy 968 gccatgtgcacgatgagttcaggacggcagcgggtggaagcccttttgcaccttgac 1027  
 Db 181 gccatgtgcacgatgagttcaggacggcagcgggtggaagcccttttgcaccttgac 239  
 Qy 1028 atgggaagacttggttacaacattccacagacagatgagtaaacctcatgaccatagcc 1087  
 Db 240 atgggaagacttggttacaacattccacagacagatgagtaaacctcatgaccatagcc 299  
 Qy 1088 tatgtcatggtgcacatcgccctctttcatgtgcacctgtgctcatggtgtgtcag 1147  
 Db 300 tatgtcatggtgcacatcgccctctttcatgtgcacctgtgctcatggtgtgtcag 359  
 Qy 1148 tggcgctgcctcgcctgctgcgcagcagcatgatgtatttgcctgtgatgacatccctcg 1207  
 Db 360 tggcgctgcctcgcctgctgcgcagcagcatgatgtatttgcctgtgatgacatccctcg 419



Fri May 12 12:27:56 2000

Search completed: May 1, 2000, 19:59:20  
Job time: 21944 sec



Query result	Score	Match	Length	DB	ID	Description
c 1	325	18.6	525	41	AI005033	AI005033 ou91b12.x









```

source      1. .428
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1689524"
/clone_lib="Soares_NHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/notes="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site: 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT 111 a 100 c 101 g 116 t
ORIGIN

Query Match      12.8%; Score 224; DB 42; Length 428;
Best Local Similarity 99.5%; Pred. No. 7.1e-98;
Matches 394; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1326 cctcccaccacccaaatgctctccctgatggagaaagcctggcaagtgsgt 1385
|||||
DB 428 CCTCCCCACCCCAAAATGCTGCTTGTATGGAGAGAAAGCGCTGGCANGTGGGT 369
|||||

QY 1386 tcacagtgactgacctgtgagagacagaaagagaaagacactctgctgcygga 1445
|||||
DB 368 TCCAGGACTGTACCTGTAGGAAACAGAAAGAGAAAGACACTCTGCTGCGGGA 309
|||||

QY 1446 atactcttggtcacctcaatttaagtcgggaattctgctgctgaaacttcagccctg 1505
|||||
DB 308 ATACTCTTGCTACCTCAAAATTAAGTCGGGAATCTGCTGTTGAACCTTCAGCCCTG 249
|||||

QY 1506 aaccttgt-caccattcttaaatctccacccaaagtattcttcttagttt 1564
|||||
DB 248 AACCTTTGCCACCATCTCTTAATCTCCACCAAGATATTCTCTTTCTTAGTTT 189
|||||

QY 1565 cagaagtactggcatcacgcaggttaccttggogtgtgctccctgtgtaccctggagc 1624
|||||
DB 188 CAGAAGTACTGGCATCACACGAGGTACCTTGGCGTGTGCTGCTGTACCTGGCAG 129
|||||

QY 1625 agaagaccaaagtgtttccctgctgccaagtacgtagagagatgcacagtgtg 1684
|||||
DB 128 AGAAGAGACCAAGCTTTTCCCTGCTGCGCAAGTCAGTAGAGAGAGATGCACAGTTTG 69
|||||

QY 1685 ctattgtcttagacagcaggactgtataaacagc 1720
|||||
DB 68 CTATTGCTTAGACAGAGGACTGTATAAACAGC 33
|||||

RESULT 8
LOCUS      AI669838      315 bp      mRNA      EST      14-MAY-1999
DEFINITION tu31h11.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2252709 3',
mRNA sequence.
ACCESSION  AI669838
VERSION     AI669838.1 GI:4834612
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 315)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL    Unpublished (1997)

```

```

COMMENT
On Dec 20, 1995 this sequence version replaced gi:1131390.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
CDNA Library Arrayed by: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.lnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco.
Location/Qualifiers
1. .315
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2252709"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr28 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615)."
Subtraction by Bento Soares and M. Fatima Bonaldo.
BASE COUNT 92 a 64 c 72 g 87 t
ORIGIN

Query Match      12.7%; Score 222; DB 50; Length 315;
Best Local Similarity 100.0%; Pred. No. 6.6e-97;
Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1515 caccattcctttaaattctccacccaaagtattcttctttttttttagtttcagaagtact 1574
|||||
DB 251 CACCATTCCCTTAAATCTCCACCCAAAGTATTCTCTTTCTTAGTTTCAGAAAGTACT 192
|||||

QY 1575 ggcatacacgcaggttaccttggcgtgtgtccctgtgtaccctggcacagaagagacc 1634
|||||
DB 191 GGCATCACACGAGGTACCTTGGCGTGTGCTGCTGGTACCTGGCAGAGAGAGACC 132
|||||

QY 1635 aagcttgttccctgctggccaaagtacgtagagagatgcacagtttgcatttgcctt 1694
|||||
DB 131 AAGCTTGTTCCTTCTGCTGGCCAAAGTCAGTAGGAGAGATGCACAGTTTGTCTATTGCTT 72
|||||

QY 1695 tagagacagggactgtataaacagcctaacattgggtgcaaa 1736
|||||
DB 71 TAGACAGAGGACTGTATAAACAGCCTTAACATTTGGTGCAAA 30
|||||

RESULT 9
LOCUS      AA705405      327 bp      mRNA      EST      24-DEC-1997
DEFINITION zj90d03.s1 Soares_fetal_liver_spleen_INFLS_S1 Homo sapiens cDNA
clone IMAGE:462149 3', mRNA sequence.
ACCESSION  AA705405
VERSION     AA705405.1 GI:2715323
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 327)
AUTHORS   Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
            Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
            Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,

```

**TITLE**  
**JOURNAL**  
**COMMENT**

Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
 WashU-NCI Human EST Project  
 Unpublished (1997)  
 On May 8, 1995 this sequence version replaced gi:800244.  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40ml3 fwd. Ef from Amersham  
 High quality sequence stop: 299.

**FEATURES**  
 source

1. 327  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3753172"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:462149"  
 /clone\_lib="Soares\_fetal\_liver\_spleen\_LNFLS\_S1"  
 /sex="male"  
 /dev\_stage="20 week-post conception fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
 with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
 This is a subcloned version of the original Soares fetal  
 liver spleen LNFLS library. 1st strand cDNA was primed  
 with a Pac I - oligo(dT) primer [5',  
 AACGGAAGAAATTAATTAAGATCTTTTTTTTTTTT 3'],  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Pac I and cloned into the Pac I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization. Library  
 constructed by Bento Soares and M.Fatima Bonaldo."

**BASE COUNT** 96 a 71 c 78 g 82 t  
**ORIGIN**

Query Match 12.7%; Score 222; DB 37; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 6.6e-97;  
 Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1515 caccattctttaaattccaaacccaagattcttttttttttagtttcagaagtact 1574  
 |||||  
 Db 238 CACCATTCTTTAAATTCCTCAACCCAAAGATTTCTTCTTAGTTTCAGAGTACT 179  
 |||||  
 Qy 1575 ggcatacacacaggttaccttggcgtgtgtccctgtgtaccctggcagagaagacc 1634  
 |||||  
 Db 178 GGCATCACACGAGGTACCTTGGCGTGTGCCCTGTGTACCCCTGGCAGAGAGACC 119  
 |||||  
 Qy 1635 aagcttggttccctgctgcccgaagtcagtaggagagatgcacagttgtctattgctt 1694  
 |||||  
 Db 118 AAGCTTGTTTCCCTGCTGCCAAAGTCAGTAGGAGAGGATGCACAGTTGCTATTGCTT 59  
 |||||  
 Qy 1695 tagagacaggagctgtataaacaagcctaaacattggtgcaaa 1736  
 |||||  
 Db 58 TAGAGACAGGAGCTGTATATAACAGCCCTAACATTTGGTGCAAA 17

**RESULT** 10  
 AI874047/c  
**LOCUS** AI874047 341 bp mRNA EST  
**DEFINITION** wm47f12.x1 NCI\_CGAP\_Ut4 Homo sapiens cDNA clone IMAGE:2439119 3',  
 mRNA sequence.  
**ACCESSION** AI874047  
**VERSION** AI874047.1 GI:5548096  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
**REFERENCE** 1 (bases 1 to 341)

**AUTHORS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On May 18, 1998 this sequence version replaced gi:3136809.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco  
 High quality sequence stop: 333.

**FEATURES**  
 source

1. 341  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2439119"  
 /clone\_lib="NCI\_CGAP\_Ut4"  
 /tissue\_type="serous papillary carcinoma, high grade, 2  
 pooled tumors"  
 /lab\_host="DH10B"  
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT;  
 Average insert size 1.48 Kb. Life Technologies catalog #:  
 11542-016"

**BASE COUNT** 96 a 75 c 78 g 92 t  
**ORIGIN**

Query Match 12.7%; Score 222; DB 62; Length 341;  
 Best Local Similarity 100.0%; Pred. No. 6.6e-97;  
 Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1515 caccattctttaaattccaaacccaagattcttttttttagtttcagaagtact 1574  
 |||||  
 Db 241 CACCATTCTTTAAATTCCTCAACCCAAAGATTTCTTCTTAGTTTCAGAGTACT 182  
 |||||  
 Qy 1575 ggcatacacacaggttaccttggcgtgtgtccctgtgtaccctggcagagaagacc 1634  
 |||||  
 Db 181 GGCATCACACGAGGTACCTTGGCGTGTGCCCTGTGTACCCCTGGCAGAGAGACC 122  
 |||||  
 Qy 1635 aagcttggttccctgctgcccgaagtcagtaggagagatgcacagttgtctattgctt 1694  
 |||||  
 Db 121 AAGCTTGTTTCCCTGCTGCCAAAGTCAGTAGGAGAGGATGCACAGTTGCTATTGCTT 62  
 |||||  
 Qy 1695 tagagacaggagctgtataaacaagcctaaacattggtgcaaa 1736  
 |||||  
 Db 61 TAGAGACAGGAGCTGTATATAACAGCCCTAACATTTGGTGCAAA 20

**RESULT** 11  
 AI081812/c  
**LOCUS** AI081812 369 bp mRNA EST  
**DEFINITION** OX77f03.x1 Soares\_NHMPu\_S1 Homo sapiens cDNA clone IMAGE:1662365  
 3', mRNA sequence.  
**ACCESSION** AI081812  
**VERSION** AI081812.1 GI:3418604  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
**REFERENCE** 1 (bases 1 to 369)  
**AUTHORS** NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index





Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco  
 High quality sequence stop: 410.

#### FEATURES

source  
 1..425  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2408056"  
 /clone\_lib="NCI\_CGAP\_Lu19"  
 /tissue\_type="squamous cell carcinoma, poorly  
 differentiated (4 pooled tumors, including primary and  
 metastatic)."  
 /dev\_stage="adult"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; 1st strand cDNA was prepared from  
 pooled lung tumor tissue, and was then primed with a Not I  
 - oligo(dT) primer. Double-stranded cDNA was ligated to  
 Eco RI adaptors (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of the modified  
 pT73 vector. Library went through one round of  
 normalization. Library constructed by Bento Soares and M.  
 Fatima Bonaldo."

BASE COUNT 109 a 99 c 89 g 128 t

Query Match 12.7%; Score 222; DB 61; Length 425;  
 Best Local Similarity 100.0%; Pred. No. 6.6e-97;  
 Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1515 caccattctttaaattctccaaacccaaagtattcttcttttagtttcagaagtact 1574  
 |||||  
 DB 258 CACCATTCTTTAAATTCCTCAACCCAAAGTATCTCTTTCTTTAGTTTCAGAGTACT 199  
 |||||  
 QY 1575 ggcatacacgaggttacttggcgtgtgtccctgtgtagcctggcagagaagagacc 1634  
 |||||  
 DB 198 GGCATCACACGAGGTACTTGGCGTGTGTCCCTGTGTGTCCTGGCAGAGAGAGACC 139  
 |||||  
 QY 1635 aagctgtttccctgtgcccgaagtcagtaggagaggtgacagttgctattgctt 1694  
 |||||  
 DB 138 AAGCTGTGTTCCCTGTGTCGCAAGTCAGTAGGAGAGATGCACAGTTTGCTATTGCTT 79  
 |||||  
 QY 1695 tagagacaggagctgtataaaacgaagcctaacttggtgcaaa 1736  
 |||||  
 DB 78 TAGAGACAGGGAGCTGTATAAACAAAGCCTAACATTGGTGCAAA 37  
 |||||

RESULT 14  
 AA677847/c 427 bp mRNA EST 19-DEC-1997  
 LOCUS z113c06.s1 Soares\_fetal\_liver\_spleen\_lnfels\_s1 Homo sapiens cDNA  
 DEFINITION clone IMAGE:430666 3', mRNA sequence.  
 AA677847  
 ACCESSION AA677847.1 GI:2658369  
 VERSION EST.  
 KEYWORDS human.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 427)  
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Krimman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
 Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
 TITLE WashU-NCI human EST Project

#### JOURNAL COMMENT

Unpublished (1997)  
 On Nov 6, 1997 this sequence version replaced gi:1565841.  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40ml3 fwd. ET from Amer-sham.

#### FEATURES

source  
 1..427  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:430666"  
 /clone\_lib="Soares\_fetal\_liver\_spleen\_lnfels\_s1"  
 /sex="male"  
 /dev\_stage="20 week-post conception fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)  
 with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
 This is a subtracted version of the original Soares fetal  
 liver spleen lNFLS library. 1st strand cDNA was primed  
 with a Pac I - oligo(dT) primer [5'  
 AACGGAGAATTAATTAAGATCTTTTTTTTTTTTTTTTTTTT 3'],  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Pac I and cloned into the Pac I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization. Library  
 constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 111 a 99 c 102 g 115 t

Query Match 12.7%; Score 222; DB 37; Length 427;  
 Best Local Similarity 100.0%; Pred. No. 6.6e-97;  
 Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1515 caccattctttaaattctccaaacccaaagtattcttcttttagtttcagaagtact 1574  
 |||||  
 DB 238 CACCATTCTTTAAATTCCTCAACCCAAAGTATCTCTTTCTTTAGTTTCAGAGTACT 179  
 |||||  
 QY 1575 ggcatacacgaggttacttggcgtgtgtccctgtgtagcctggcagagaagagacc 1634  
 |||||  
 DB 178 GGCATCACACGAGGTACTTGGCGTGTGTCCCTGTGTGTCCTGGCAGAGAGAGACC 119  
 |||||  
 QY 1635 aagctgtttccctgtgcccgaagtcagtaggagaggtgacagttgctattgctt 1694  
 |||||  
 DB 118 AAGCTGTGTTCCCTGTGTCGCAAGTCAGTAGGAGAGATGCACAGTTTGCTATTGCTT 59  
 |||||  
 QY 1695 tagagacaggagctgtataaaacgaagcctaacttggtgcaaa 1736  
 |||||  
 DB 58 TAGAGACAGGGAGCTGTATAAACAAAGCCTAACATTGGTGCAAA 17  
 |||||

RESULT 15  
 AA136283/c 434 bp mRNA EST 30-NOV-1996  
 LOCUS zK93d05.s1 Soares\_pregnant\_uterus\_NbHPU Homo sapiens cDNA clone  
 DEFINITION IMAGE:490377 3', mRNA sequence.  
 AA136283  
 ACCESSION AA136283  
 VERSION EST.  
 KEYWORDS human.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 434)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and

Search completed: May 1, 2000, 14:44:09  
Job time: 13969 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:07:05 ; Search time 4425.31 Seconds  
(without alignments)  
-959.224 Million cell updates/sec

Title: us-09-215-435-71

Perfect score: 1398

Sequence: 1 gatgcctgagggcccgagc.....acatttgaaaaaaaaaaaaa 1398

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 821193 seqs, -1518192014 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba1.\*

2: gb\_ba2.\*

3: gb\_om.\*

4: gb\_ov.\*

5: gb\_pat.\*

6: gb\_ph.\*

7: gb\_pil.\*

8: gb\_pl2.\*

9: gb\_pr1.\*

10: gb\_pr2.\*

11: gb\_pr3.\*

12: gb\_ro.\*

13: gb\_sts.\*

14: gb\_sy.\*

15: gb\_un.\*

16: gb\_v1.\*

17: em\_fun.\*

18: em\_hum1.\*

19: em\_hum2.\*

20: em\_in.\*

21: em\_om.\*

22: em\_or.\*

23: em\_lov.\*

24: em\_pat.\*

25: em\_ph.\*

26: em\_pl.\*

27: em\_ro.\*

28: em\_sts.\*

29: em\_sy.\*

30: em\_un.\*

31: em\_v1.\*

32: gb\_htg1.\*

33: gb\_htg2.\*

34: gb\_in1.\*

35: gb\_in2.\*

36: em\_ba1.\*

37: em\_ba2.\*

38: em\_hum3.\*

39: em\_hum4.\*

40: gb\_pr4.\*

41: gb\_htg3.\*

42: gb\_htg4.\*

43: gb\_htg5.\*

44: gb\_htg6.\*  
45: gb\_htg7.\*  
46: em\_htg1.\*  
47: em\_htg2.\*  
48: em\_htg3.\*  
49: em\_hum5.\*  
50: gb\_pl3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	176	12.6	461	13	G13307	G13307 human STS W
c 2	23	1.6	104634	41	AC011384	AC011384 Homo sapi
c 3	22	1.6	9746	35	AE001415	AE001415 Plasmodiu
c 4	22	1.6	92611	8	ATT4D2	AL132958 Arabidops
c 5	22	1.6	153098	34	PFMAL3P2	AL034558 Plasmodiu
c 6	22	1.6	156420	40	AC007510	AC007510 Homo sapi
c 7	21	1.5	1182	34	AF021905	AF021905 Caenorhab
c 8	21	1.5	1267	4	RPAJ2554	AJ002554 Rana pere
c 9	21	1.5	3258	12	MMU70324	U70324 Mus musculu
c 10	21	1.5	3323	12	AB025259	AB025259 Mus muscu
c 11	21	1.5	3741	12	AB025258	AB025258 Mus muscu
c 12	21	1.5	41378	11	AC005381	AC005381 Homo sapi
c 13	21	1.5	160969	44	AC016692	AC016692 Homo sapi
c 14	21	1.5	177707	40	AC006265	AC006265 Homo sapi
c 15	21	1.5	178524	40	AC006427	AC006427 Homo sapi
c 16	20	1.4	289	13	G45108	G45108 Dictyostell
c 17	20	1.4	389	34	DB438	X53444 Dictyostell
c 18	20	1.4	863	5	I24979	I24979 Sequence 11
c 19	20	1.4	863	5	I92698	I92698 Sequence 11
c 20	20	1.4	1278	4	AF128814	AF128814 Oryzias l
c 21	20	1.4	1479	4	GFVILIP	AF181652 Drosophil
c 22	20	1.4	4264	35	AF181652	AF181652 Drosophil
c 23	20	1.4	5618	9	AB007862	AB007862 Homo sapi
c 24	20	1.4	6594	7	IPBSTAPHOA	L25626 Ipomoea bat
c 25	20	1.4	6788	9	HUMNYC	M13241 Human N-myc
c 26	20	1.4	8762	9	HSNMYC	Y00664 Human germ
c 27	20	1.4	10436	40	HSU52962	U52962 Human cent
c 28	20	1.4	15975	12	D10911	D10911 Mus musculu
c 29	20	1.4	36277	10	AP000336	AP000336 Homo sapi
c 30	20	1.4	37427	34	CEH12119	Z98851 Caenorhabdi
c 31	20	1.4	38750	41	AC009081	AC009081 Homo sapi
c 32	20	1.4	62636	43	AC015634	AC015634 Homo sapi
c 33	20	1.4	64326	10	AP000331	AP000331 Homo sapi
c 34	20	1.4	88037	32	PFMAL13P8	AL036782 Plasmodiu
c 35	20	1.4	89820	40	AC007868	AC007868 Genomic S
c 36	20	1.4	100000	10	AP000172	AP000172 Homo sapi
c 37	20	1.4	100000	10	AP000215	AP000215 Homo sapi
c 38	20	1.4	100000	10	AP000504	AP000504 Homo sapi
c 39	20	1.4	107300	33	AC007345	AC007345 Homo sapi
c 40	20	1.4	107897	10	AP000057	AP000057 Homo sapi
c 41	20	1.4	112235	40	HUAC004685	AC004685 Homo sapi
c 42	20	1.4	117199	10	AP000125	AP000125 Homo sapi
c 43	20	1.4	119136	11	AC004222	AC004222 Homo sapi
c 44	20	1.4	120060	41	AC005505	AC005505 Plasmodiu
c 45	20	1.4	135130	41	AC011096	AC011096 Homo sapi

#### ALIGNMENTS

RESULT 1  
G13307/c G13307 461 bp DNA STS  
LOCUS human STS WI-12510, sequence tagged site.  
ACCESSION G13307  
VERSION G13307.1 GI:1127416

04-JUN-1996

RESULT	2
AC011384/c	
LOCUS	
DEFINITION	AC011384 104634 bp DNA HTG 06-OCT-1999 Homo sapiens chromosome 5 clone CIT978SKB_143D11, *** SEQUENCING IN PROGRESS ***, 5 ordered pieces.
ACCESSION	AC011384
VERSION	AC011384.1 GI:6013557
KEYWORDS	HTG; HTGS_PHASE2.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. DOE Joint Genome Institute. Sequencing of Human Chromosome 5 Unpublished 2.(bases 1 to 104634) DOE Joint genome institute. Direct Submission Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA www.jgi.doe.gov. * NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter. * This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved * 1 29369: contig of 29369 bp in length * 29370 36782: gap of unknown length * 36783 51002: contig of 14220 bp in length * 51003 77297: contig of 26295 bp in length * 77298 104634: contig of 27337 bp in length. Location/Qualifiers 1..104634 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosomes="5" /clone="CIT978SKB_143D11"
FEATURES	
source	
BASE COUNT	31993 a 19210 c 19766 g 33664 t 1 others
ORIGIN	
Query Match	1.6%; Score 23; DB 41; Length 104634;
Best Local Similarity	100.0%; Pred.No.0.11;
Matches 23; Conservative	0; Mismatches 0; Indels 0; Gaps 0;

Db	6357	GATACATTGAAAAA	6335
RESULT	3		
AE001415/c			
LOCUS	AE001415	9746 bp	DNA
DEFINITION	Plasmodium falciparum chromosome 2, section 52 of 73 of the complete sequence.		
VERSION	AE001415	AE001362	
ACCESSION	AE001415.1	GI:3845264	
KEYWORDS	malaria parasite P. falciparum.		
SOURCE	Plasmodium falciparum		
ORGANISM	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
REFERENCE	1 (bases 1 to 9746)		
AUTHORS	Gardner,M.J., Tettelin,H., Carucci,D.J., Cummings,L.M., Aravind,L., Koonin,E.V., Shalton,S., Mason,T., Yu,K., Fujii,C., Pederson,J.,		

Shen, K., Jing, J., Aston, C., Lai, Z., Schwartz, D.C., Pertea, M., Salzberg, S., Zhou, L., Sutton, G.G., Clayton, R., White, O., Smith, H.O., Fraser, C.M., Adams, M.D., Venter, J.C. and Hoffman, S.L. Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum

JOURNAL  
MEDLINE  
REMARK  
Science 282 (5391), 1126-1132 (1998)  
99021743  
Erratum: [[published erratum appears in Science 1998 Dec 4; 282(5395):1827]]  
2 (bases 1 to 9746)  
Gardner, M.J.  
Direct Submission  
Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA  
JOURNAL  
TITLE  
Location/Qualifiers  
source  
1. 9746  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:5833"  
/chromosome="2"  
complement(365..1006)  
/gene="PF0740c"  
complement(365..1006)  
/gene="PF0740c"  
/note="Predicted by GlimmerM"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAC71937.1"  
/db\_xref="GI:3845265"  
/translation="MNNKKKKKTHNVIENIOKCSFVLKKNNDIYSSNNKPIQI  
YNDRIKLAEHTKNEVEELLEGDYKDLNKNKYINDAVYIHAVGINLLKASYIQLDF  
SYHEFKVSGEPIISNNNNNNNILEKKKKKKKNPLAYIDIHTECHTLNWDNI  
ITNIYDQDHFNNHNNDDKDSFYDEYDNLKIFASMYDPLKHKYLEVRKKK"  
1578..5722  
/gene="PF0745w"  
join(1578..4550,4640..5722)  
/gene="PF0745w"  
/note="Predicted by GlimmerM"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAC71938.1"  
/db\_xref="GI:3845266"  
/translation="MFVYVONISNRINKYKINLTCSRYNKIFCLHKSSISSK  
ANINIRFVFTKQYDERVKLLKVLREKDFELISKKNKKECVSNNIYNN  
NNINLIYDSTKYGNGDDAAIEKSPPLNNEKNKKNKILLYNKKIKLIDKK  
CNIISILNKKSYFTLSCVNIIRNDIFNYFLKCLYNOWHILNTNMVSLFL  
NVSTLYCEEKINRYTHRRHYHILVNFLEICNVCYKNIQVLPVLLICDK  
KLDSTLGFNLKIIIMFFKTHRNALLVTHSTEEELIHKRISFLYKNNRNNIQ  
HDDINNETDNVKNLYGRKKKNLYGNNNNNNNNNKNNKKSISTNLLNKKYKNEH  
IVTHVINKDEKKELFCFTFVNMTLLIYELFYKNISTNNKINTEYIDDTWNNII  
TNIIYTKNNIPMERIKKETHQLSIIISLLYSLVINYSKLYENIFYIFERSVDIIHL  
FKNNKINIMTFDELKNDLVNFVNCDDNNNNNDNNDDNNDDNNDDNNDDNN  
DNNVIKYKHSNPEPKYKYNKYNNNTFHRNIFKYQNIYVHNLAKIDPLLYNNEL  
FVYVDDLYSDNCTDMFTDLCKLVALSYQKEIEKOKKKKKRYHFKDIISL  
LPYNTIVEROIFLLYNNKNNICKIETCNLYNNYDVPVYKNNKLVAGWEKN  
NYDKNSILSSYKNLYNCNDNTYVHSSYICIEKLSHLLNIIYQHKIVDKMFIY  
LTFLLMPKKYKIIIDILFNIINALSRCMYTEMVVLIFYVNVCGIRISEYLSKY  
FTRGLVLLVYEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE  
EKDEKEEKEEIOIKKQYKVKKNKKKNKKNKKNKKNKKNKKNKKNKKNKKNK  
LNKQFNKQDQHLNLSLVPYVCKGSEVNFRRNNYNNNNSEKIDVYNNYK  
KNNKVKYISSNDKMEKEKEFNICLNTLLANNMVTNDIOYNNKVLHKIKNT  
IKKEKIDVRLSISIFSYRNLNIIYKLYNNYKQLOKLNFGNISILSYMNT  
AIYDKHIDVLCCKDIFKNDKNNIQQNQLSHLHFLFILTSLFOLFNFKFDVLSY  
IFRILYIYIYVNNQIITIKKKKNQSFQYNNINISSVITTPPKNFISMFDISNL  
YHFFLLPLPHNNKVICVNNISHLNLSLSYKYYKHYHVAPTSPDIQRQVLNVNK  
MLLGHGNIKVSVEYKMHNNPQIDILIIKYG"  
6595..8763  
/gene="PF0750w"  
/gene="PF0750w"  
/note="Identified by sequence similarity; putative"  
/codon\_start=1  
/product="VPS45-like protein (STXBP/UNC-18/SECI family)"  
/protein\_id="AAC71939.1"

Shen, K., Jing, J., Aston, C., Lai, Z., Schwartz, D.C., Pertea, M., Salzberg, S., Zhou, L., Sutton, G.G., Clayton, R., White, O., Smith, H.O., Fraser, C.M., Adams, M.D., Venter, J.C. and Hoffman, S.L. Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum

JOURNAL  
MEDLINE  
REMARK  
Science 282 (5391), 1126-1132 (1998)  
99021743  
Erratum: [[published erratum appears in Science 1998 Dec 4; 282(5395):1827]]  
2 (bases 1 to 9746)  
Gardner, M.J.  
Direct Submission  
Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA  
JOURNAL  
TITLE  
Location/Qualifiers  
source  
1. 9746  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:5833"  
/chromosome="2"  
complement(365..1006)  
/gene="PF0740c"  
complement(365..1006)  
/gene="PF0740c"  
/note="Predicted by GlimmerM"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAC71937.1"  
/db\_xref="GI:3845265"  
/translation="MNNKKKKKTHNVIENIOKCSFVLKKNNDIYSSNNKPIQI  
YNDRIKLAEHTKNEVEELLEGDYKDLNKNKYINDAVYIHAVGINLLKASYIQLDF  
SYHEFKVSGEPIISNNNNNNNILEKKKKKKKNPLAYIDIHTECHTLNWDNI  
ITNIYDQDHFNNHNNDDKDSFYDEYDNLKIFASMYDPLKHKYLEVRKKK"  
1578..5722  
/gene="PF0745w"  
join(1578..4550,4640..5722)  
/gene="PF0745w"  
/note="Predicted by GlimmerM"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAC71938.1"  
/db\_xref="GI:3845266"  
/translation="MFVYVONISNRINKYKINLTCSRYNKIFCLHKSSISSK  
ANINIRFVFTKQYDERVKLLKVLREKDFELISKKNKKECVSNNIYNN  
NNINLIYDSTKYGNGDDAAIEKSPPLNNEKNKKNKILLYNKKIKLIDKK  
CNIISILNKKSYFTLSCVNIIRNDIFNYFLKCLYNOWHILNTNMVSLFL  
NVSTLYCEEKINRYTHRRHYHILVNFLEICNVCYKNIQVLPVLLICDK  
KLDSTLGFNLKIIIMFFKTHRNALLVTHSTEEELIHKRISFLYKNNRNNIQ  
HDDINNETDNVKNLYGRKKKNLYGNNNNNNNNNKNNKKSISTNLLNKKYKNEH  
IVTHVINKDEKKELFCFTFVNMTLLIYELFYKNISTNNKINTEYIDDTWNNII  
TNIIYTKNNIPMERIKKETHQLSIIISLLYSLVINYSKLYENIFYIFERSVDIIHL  
FKNNKINIMTFDELKNDLVNFVNCDDNNNNNDNNDDNNDDNNDDNNDDNN  
DNNVIKYKHSNPEPKYKYNKYNNNTFHRNIFKYQNIYVHNLAKIDPLLYNNEL  
FVYVDDLYSDNCTDMFTDLCKLVALSYQKEIEKOKKKKKRYHFKDIISL  
LPYNTIVEROIFLLYNNKNNICKIETCNLYNNYDVPVYKNNKLVAGWEKN  
NYDKNSILSSYKNLYNCNDNTYVHSSYICIEKLSHLLNIIYQHKIVDKMFIY  
LTFLLMPKKYKIIIDILFNIINALSRCMYTEMVVLIFYVNVCGIRISEYLSKY  
FTRGLVLLVYEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE  
EKDEKEEKEEIOIKKQYKVKKNKKKNKKNKKNKKNKKNKKNKKNKKNKKNK  
LNKQFNKQDQHLNLSLVPYVCKGSEVNFRRNNYNNNNSEKIDVYNNYK  
KNNKVKYISSNDKMEKEKEFNICLNTLLANNMVTNDIOYNNKVLHKIKNT  
IKKEKIDVRLSISIFSYRNLNIIYKLYNNYKQLOKLNFGNISILSYMNT  
AIYDKHIDVLCCKDIFKNDKNNIQQNQLSHLHFLFILTSLFOLFNFKFDVLSY  
IFRILYIYIYVNNQIITIKKKKNQSFQYNNINISSVITTPPKNFISMFDISNL  
YHFFLLPLPHNNKVICVNNISHLNLSLSYKYYKHYHVAPTSPDIQRQVLNVNK  
MLLGHGNIKVSVEYKMHNNPQIDILIIKYG"  
6595..8763  
/gene="PF0750w"  
/gene="PF0750w"  
/note="Identified by sequence similarity; putative"  
/codon\_start=1  
/product="VPS45-like protein (STXBP/UNC-18/SECI family)"  
/protein\_id="AAC71939.1"

/db\_xref="GI:3845267"  
/translation="MENNPFVFKSLVNIYEEYLNLIHRVKGKVLVDDETSIISL  
IFGHSYLEKEIFLTFNFDKNIFDIYNNNDKKNFDMFKYKIKNKLKLVIFLLR  
PTVNTILRLMSLKLFLSEYIYFFNTINDIYIEKLAKADEFDVKNIIYEYIDTV  
LHDYLFHNDIYTSFLYKNDHKFDKREKKELNYFKOYNNNNNNNNNNSSDGRYE  
KLIEBPNKLEGNMIIYDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
LYMNQIVQRIIDGLFSFCICRQVPOVYINRHSKICHIIDMLKMLRQSQFNNIL  
DIYEKYNDENRERKKKKKILETNNEPNYQFNHLINQNIHEITGDACYFLIDRNEP  
TPLTQWYQSMLEHIGIENNNINLNKNKEEQOQIVMNSCNYVHEHLEDFNFDG  
LQGVKNYVDIYQEBTSKTNLESIDDIQFIDIPYNNKLSGNTKRVNLHFKFSQ  
VQKROLFYISELQSIACVHTKNDHFQVIDTIKNYTYTNDVLRLLSLYSLKADQ  
HINVINELAKRNIQKQDILLIDALLYSQOQKYNQLEKQFELAKTINPTING  
TSNFTLHKSILYILEDDIKYIKNTQIYNTLLHTEPINKKINSIVVFFIGAYI  
EYRDYQHLUSKYNISIVLGSSTHMSQSFLADVLQLIK"  
BASE COUNT 4253 a 831 c 1030 g 3632 t  
ORIGIN

Query Match 1.6%; Score 22; DB 35; Length 9746;  
Best Local Similarity 100.0%; Pred. No. 0.37;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1377 ataccattgaaaaa 1398  
|||||  
Db 8963 ATACATTGAAAAA 8942  
|||||

RESULT 4  
ATT4D2 92611 bp DNA PLN 12-NOV-1999  
LOCUS Arabidopsis thaliana DNA chromosome 3, BAC clone T4D2.  
DEFINITION AL132958  
ACCESSION AL132958  
VERSION AL132958.1 GI:6434245  
KEYWORDS  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
Arabidopsis.  
REFERENCE 1 (bases 1 to 92611)  
AUTHORS Nyakatura, G., Partmann, B., Dauner, D., Sterr, W., Holland, R.,  
Weichselgartner, M., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quetier, F.,  
and Salanoubat, M.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 92611)  
AUTHORS EU Arabidopsis sequencing project.  
TITLE Direct Submission  
JOURNAL Submitted (12-NOV-1999) MIPS, at the Max-Planck-Institut fuer  
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:  
lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project  
Coordinator: Marcel Salanoubat and Francis Quetier, Groupement  
d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue  
Gaston Cremieux, BP191, 91006 Evry Cedex, France;  
http://www.genoscope.cns.fr  
COMMENT Information on performance of analysis and a more detailed  
annotation of this entry and other sequences of chromosomes 3, 4  
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.  
FEATURES  
source  
1. 92611  
/organism="Arabidopsis thaliana"  
/variety="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="3"  
BASE COUNT 29024 a 16857 c 16763 g 29967 t  
ORIGIN

Query Match 1.6%; Score 22; DB 8; Length 92611;  
Best Local Similarity 100.0%; Pred. No. 0.4;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1377 ataccattgaaaaa 1398

|||||  
Db 79365 ATACATTGAAAAA 79386

RESULT 5  
PFMAL3P2 153098 bp DNA INV 28-JUL-1999  
LOCUS Plasmodium falciparum MAL3P2, complete sequence.  
DEFINITION AL034558 AL008982 AL009007 AL009008 AL009009 AL009014  
ACCESSION AL010140 AL010150 AL010158 AL010211 AL022217 Z97350 Z98546  
VERSION AL034558.2 GI:4493878  
KEYWORDS HTG  
SOURCE malaria parasite P. falciparum.  
ORGANISM Plasmodium falciparum  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 153098)  
AUTHORS Churcher,C., Bowman,S., Lawson,S., Quail,M. and Barrell,B.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 153098)  
AUTHORS Lawson,D., Bowman,S. and Barrell,B.  
TITLE Direct Submission  
JOURNAL Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium,  
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA, UK  
COMMENT On Mar 24, 1999 this sequence version replaced gi:4034874.  
For more information about this sequence or the Malaria Project,  
see <http://www.sanger.ac.uk/projects/p-falciparum>. IMPORTANT: This  
sequence is not the entire insert of clone MAL3P2. It may be  
shorter because we only sequence overlapping sections once, or  
longer because we arrange for a small overlap between neighbouring  
submissions.

FEATURES

Source  
1. .153098  
/organism="Plasmodium falciparum"  
/strain="3D7"  
/db\_xref="taxon:5833"  
/chromosome="3"  
/clone="MAL3P2"  
1030. .2484  
/gene="PFC0160w"  
1030. .2484  
/gene="PFC0160w"  
/note="predicted using hexExon: MAL3P2.1 (PFC0160w),  
Putative Ankyrin repeat protein, len: 285 aa"  
/codon\_start=1  
/protein\_id="CAB38988.1"  
/db\_xref="GI:4493879"  
/db\_xref="SPTREMBL:O97224"  
/translation="MNDSELDSEVLEKQYELIKYAKYQDFIRLOILQIPLVLLNDIE  
MNSINILHWACYCGFTLVKVLKLSFNCDEKEDLVNNDTPYIYAKNSVEIVLLLI  
KCFGSLFHKWRMSPLTAICEFNEDKILDEALHILELLTNGVSLSEEQNEHQTA  
LFLSVKNNTSTQLMQLTKEVINHRDFGVTLVIAVRHCDIDILRLCDYGLWV  
YSSSENKNTYFQICIKNRFLYLLKRWYQKICSLKICTIYAFYEFVAIL  
NLIYVNTAHSEFSLIKYHFKSLIWTIWFQOQLKMLYFKSPGYKFNHFNKVK  
NTOSKYNPNWNGTFTNAEYOLNIETEIFOINKLSTNFTLTPINHDOALS  
NKYDILNLEYKILSYQVQSRINSDEYRALLYNQPRVKSHEKLYIYSTVL  
HIYIYIFPFFFEYLCDFLQYIKSGSLC"  
5344. .13329  
/gene="PFC0165w"  
Join(5344. .12004,12737. .13329)  
/gene="PFC0165w"  
/note="predicted using hexExon: MAL3P2.2 (PFC0165w),  
Hypothetical protein, len: 1676 aa"  
/codon\_start=1  
/protein\_id="CAB38989.1"  
/db\_xref="GI:4493880"  
/db\_xref="SPTREMBL:O97225"  
/translation="MNRKLIKLEKNRMNKNKVRDCHCMNNYSQCYVNVKN  
KYENDFKYENIYGNNDLIKRSVQEEKKKWNEDNKNKSSLEFTPYRESNRNI  
FLIENSNLNGSCVYNNDLIYIHNKRYEHHMLPKRKKKKKKKKKKKKKKRKK  
RFLHKNNDTFSVTSMDKEDYKQYDHDKDDDDDDDDDDDDDDDDDDDDDD  
EDDDDDDDDDDDNNYSYDNDNQYNTKHNKKINPNYNTIFETSNMFCVADKEK  
RKKKEQKQNSIDKRNIIYSDSDNNYEHIF"HINSDLFFLSIDSHIEKENNLIYQN

gene  
CDS

gene  
CDS

BOFINVDVIFRKFPHKFLFVMDSCNSWIVIKKSLILNEYITRQKNDLVDINDVL  
YDVNMHYRIYCYTRYIINDDFLINNKTKELIGVQVASHINTNTNNMNDIHI  
NNNNINNMREHLKRIEKKKNKNNNNNNNDDELSTDTGDMSLNRCNEES  
KYKVLVLKMTLAKHLFNQDINILFHFNACLSNYNKLDENTSYLLNNYNLNKS  
EHLNISRKNTYTLDTHTNTYNSNTIITKNYSNNNNNNNDNLNDLYSYFNNY  
LTKKIVENYSYSLYIKECLLIIDLILYNTYGNLIYIKYYKVDHENTYNIISNI  
VOKKNEAKKSKYNIYHSEKNNRKQSNYSISSFSQKNYKNTNNNNLHGE  
IPDSMSSSNTSSSNSSSNSSNSATSYPDKMRNRKNLNYINKQNDNDLIYF  
FPYKLIILPTLNVNEQNSNIEARILYIGYIRKIQFIEIKNSQSCNFTDI  
FCCCLSNISAILHNHINRIENVRNIYFTKINMKHSTNNNNNNNNNDNNFIH  
NDDLSLCLDEVKLSCTDPEEKLFPSINLHKNIODENLFEINTEKAEADISLYK  
IKISVPLIQIGKLLNKILKIKKKNNSKYLIDLINLQEIYLPDIDINKKVLILY  
LKNITPLDFIKNYFYGVKVDTYKEFFIHENKHITPYIKONKHPYNNMYKLFKK  
YIDLSTNYGARYIILNDKVPKFLKSNHIFITGKYIDVLSAKSLINKRKTSFYL  
IDQYKYNINENIDNDTQNNRNSFLPKDVIYKRIYENINNDQYTNDSHINRNY  
FYNDLYKSEOTQGMKQIINGKVYNYMLPYEDEYEHKNNDKKNKSLYIDND  
KKVYENLIHQHLYASKWFFELYIKKIDIKKIKHHYFFELQISLDLYKYFLYSHEY  
LEKLYHKKNLILAKLNFDDISRSSVLYHLKYKKNYVAISDIFSMDVNLIDL  
RKQISNEQQLKNGANNYEEVHDGDDDDYYNNNGEYNNYEEGYNDDEERNYL  
NKNDYDDEADEDEEDDDDDYDNDNDDDDKEYEYDNGDNDYDEADSDHNNYNTI  
ENYHDDQVHNSFNINKPLHMDNTNQKQYNEEYKTKNERIKKNAELKNGAI  
KKKKITKKKNVSKLNELRKNKNTNNMNTNNNNNNNNNNNDGSGNPKNIKN  
NFILKIHNNVPNSMNVYIKGLISYKNMPEYNLIFNITIFKYTILFVPLNYCKY  
IEHKLSEVLMNMFVKNINEECTNNLMICIHRECMHFKLCYLIHQNDYKSEY  
NNMTKLAKETLIDFDIIHNTYLNILKYSIFMDNNTIILKILISLISHFTRILK  
FNFDKQEIENITHNGVSKVNLAPNTTHINHPHNNNNNNNNNNNNYHKKIIOEL  
LRDAYISMIQNTIKYDKHFKFLLTEYVNNNIDYAHNFKLIKIDYFNYYTKY  
LSYQPTFNNDINSEYVNDNDLSDNDSTROVTKENGVNTNNNNVKNHMQTS  
NNPTESVQIKRKKLLMDNRYNNDLNHNTINDMNIINLKNKGENLYKRENI  
TSEKLNHSPNEYRNLIHTHTNNNGSNYHOLNYSLHFSNNNNNNNNNNKVSOLH  
NLNKSNEKKDYNINDTNNNNHYTNNTPSKSSNEKNKIYNHNPINMYKRDYKN  
ILNNKDOTILSENPLYKKEKHIPFQKNKHNDPSPHUHNNFODENNNISPLNYSKLLK  
FNTSYNNENKINLSNIDSDSDNSYSIYIYIYIYIYLILFFYMAQOVNDPPK  
TYPNDKTAEYKAYANNMYLYYQNNELKIDSSYFKDKLGLGFPFGSWCYVTFID  
SLNIFKKNPNVEIYIPDRTYQYQSLKNTNFYALPFDNYLYICKYQIKNLPSF  
MLJTPNNNLYKDAQLIKTDEYINLKLKNIYIHKPTQFNRRFDLFRN"  
complement(13855. .15201)  
/gene="PFC0170c"  
complement(13855. .15201)  
/gene="PFC0170c"  
/note="predicted using hexExon: MAL3P2.3 (PFC0170c)  
Lipoamide acyltransferase, len: 449 aa: Similarity to  
lipoamide acyltransferases. C.elegans lipoamide  
acyltransferase (WP:2K669.4) BLAST Score: 582, sum P(1) =  
9.2e-57; 33% identity in 430 aa overlap."  
/codon\_start=1  
/protein\_id="CAB38991.1"  
/db\_xref="GI:4493882"  
/db\_xref="SPTREMBL:O97227"  
/translation="MFVKVNLNLRLEGKSFGRHYLNTSAIHFKYVKCLFDIGEG  
ISEVEITKWKNEGQVSEMSLTLTQSDKAADVITSYNGVLYVKKYLINENDMLKVS  
YFEITDDDDIERDEEVEKEBENKKEDESDLSKNDISNDYIKASPGVKRRAK  
FYKVLNKGVDYFNKNISLEDLELYNNVYNNYNNNNNDMDIIEVSLCFLKLA  
MCKSNESLOVPLFLHNMCIINLILKMKVEYKQOKNLOTKETNITITCLIKLISN  
VLKEPILNSKFNKNTYTKYKHNISIAVDTPHGLVLPNKVKNONKIDIDKDL  
SERKANNNQDLSDINGTITVSNFGAISGTFATPIYVDFDQACILGIGKMEKKLLK  
DESNLNNQDLILVADITINFTFGADHYIDGATLAQFSKMLKMNIENTCASGLPDLLE"  
17040. .18821  
/gene="PFC0175w"  
17040. .18821  
/gene="PFC0175w"  
/note="predicted using hexExon: MAL3P2.4 (PFC0175w),  
Hypothetical protein, len: 594 aa: Similarity with a  
number of hypothetical proteins from model organisms.  
C.elegans hypothetical protein (TR:O01963) BLAST Score:  
361, sum P(1) = 4.0e-47, 35% identity in 227 aa overlap."  
/codon\_start=1  
/protein\_id="CAB38990.1"  
/db\_xref="GI:4493881"  
/db\_xref="SPTREMBL:O97226"  
/translation="MNIPKEQNKNDQINVTIEYGRKMMKVIKTPINDTKNNIINH  
YDSINDKYNTSDCYLMNEDGILNLYNKKFLFOENNLNLYVELADIEKME  
LAILEIMNLINIQDYAHFELKAYNFNSNDLENWFNNSKKVLTKLKLSHLEEDILN



```

repeat_region /rpt_family="MIR"
3103..3141
repeat_region /rpt_family="(TA)n"
3239..3286
repeat_region /rpt_family="(TA)n"
complement(3404..3829)
/rpt_family="LIME1"
3830..4003
repeat_region /rpt_family="(TA)n"
complement(4004..4447)
/rpt_family="LIME1"
4452..4528
repeat_region /rpt_family="T-rich"
complement(4624..4923)
/rpt_family="AluJo"
repeat_region /rpt_family="LIME1"
complement(4927..5032)
/rpt_family="LIME1"
repeat_region complement(5034..5161)
/rpt_family="LIMEC"
5247..5309
repeat_region /rpt_family="LIMEC"
5310..5331
repeat_region /rpt_family="(TA)n"
complement(5524..5696)
/rpt_family="MIR"
repeat_region complement(6627..7180)
/rpt_family="MER41A"
repeat_region complement(8503..8803)
/rpt_family="AluSp"
repeat_region complement(9433..9695)
/rpt_family="MER47"
repeat_region complement(9707..9916)
/rpt_family="MER47"
repeat_region complement(10319..10610)
/rpt_family="AluSx"
10910..11030
repeat_region /rpt_family="MIR"
11835..12203
repeat_region 13455..13849
/rpt_family="MLT1"
complement(14148..14450)
/rpt_family="AluSc"
repeat_region complement(14682..14865)
/rpt_family="MER5A"
16744..16768
repeat_region /rpt_family="(TC)n"
17445..17499
/rpt_family="MADE1"
17518..17568
repeat_region /rpt_family="MADE1"
complement(17657..17841)
/rpt_family="LIME7"
repeat_region complement(19299..19368)
/rpt_family="MIR"
21047..21083
repeat_region /rpt_family="(TCCC)n"
23618..23687
/rpt_family="(CCCCG)n"
complement(24454..24552)
/rpt_family="MER5A"
28441..28483
repeat_region /rpt_family="L2"
31640..31936
repeat_region /rpt_family="AluY"
33528..33553
/rpt_family="(TA)n"
complement(33760..33965)
/rpt_family="MIR"
repeat_region complement(34173..35044)
/rpt_family="LIME4"
complement(35045..35355)
/rpt_family="AluYb8"

```

```

repeat_region complement(35356..35517)
/rpt_family="LIME4"
35882..36234
/rpt_family="MLT1G1"
repeat_region complement(36384..37011)
/rpt_family="LIME1"
repeat_region complement(37012..37315)
/rpt_family="AluSg1"
repeat_region complement(37316..37998)
/rpt_family="LIME1"
repeat_region complement(38006..38561)
/rpt_family="LIME4"
repeat_region complement(38567..38717)
/rpt_family="LIMEC"
repeat_region complement(39483..39759)
/rpt_family="LIMEC"
repeat_region complement(39929..40091)
/rpt_family="LIME4"
repeat_region complement(40909..40997)
/rpt_family="MIR"
repeat_region complement(41203..41317)
/rpt_family="MER103"
repeat_region complement(41342..41813)
/rpt_family="LTR33"
42498..42561
repeat_region /rpt_family="(TG)n"
42753..43035
repeat_region /rpt_family="AluSx"
43036..43078
repeat_region /rpt_family="(CAAA)n"
complement(43238..43411)
/rpt_family="MER5A"
repeat_region 43634..43914
/rpt_family="AluYb8"
43915..43944
repeat_region /rpt_family="(CAA)n"
45972..46076
repeat_region /rpt_family="(TA)n"
46077..46108
repeat_region /rpt_family="(CA)n"
complement(46136..46330)
/rpt_family="MIR"
repeat_region complement(46833..47146)
/rpt_family="AluSc"
47241..47401
repeat_region /rpt_family="L2"
47687..48024
repeat_region /rpt_family="AluSg"
48104..48406
repeat_region /rpt_family="AluY"
49230..49254
repeat_region /rpt_family="(A)n"
49476..49520
repeat_region

```

Query Match 1.6%; Score 22; DB 40; Length 156420;  
 Best Local Similarity 100.0%; Pred. No. 0.41; Mismatches 0; Indels 0; Gaps 0;

QY 1377 atacattgaaaaa 1398  
 Db 8284 ATACATTGAAAAA 8263

RESULT 7  
 AF021905  
 LOCUS AF021905 1182 bp DNA INV 15-NOV-1997  
 DEFINITION Caenorhabditis elegans presentilin (HOP-1) gene, complete cds.  
 ACCESSION AF021905  
 VERSION AF021905.1 GI:2618986  
 KEYWORDS  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;



REFERENCE 1 (bases 1 to 1182)  
 AUTHORS Li.X. and Greenwald,I.  
 TITLE HOP-1, a caenorhabditis elegans presentinlin, appears to be functionally redundant with SEL-12 presentinlin and to facilitate LIN-12 and GUP-1 signaling  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 94 (22), 12204-12209 (1997)  
 MEDLINE 98004548  
 REFERENCE 2 (bases 1 to 1182)  
 AUTHORS Li.X. and Greenwald,I.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-AUG-1997) Biochemistry, Columbia University, 701 West 168th Street, New York, NY 10032, USA  
 FEATURES  
 source  
 1. .1182  
 /organism="Caenorhabditis elegans"  
 /strain="Bristol"  
 /db\_xref="taxon:6239"  
 /chromosome="1"  
 /note="cosmid C18E3"  
 2. .1078  
 /gene="HOP-1"  
 2. .1078  
 /gene="HOP-1"  
 /note="splice leader SL1 is trans-spliced to the first nucleotide in mRNA"  
 /codon\_start=1  
 /product="presentinlin"  
 /protein\_id="AAB84394.1"  
 /db\_xref="GI:2618987"  
 /translation="MPRTKRVYSGKITGVLYPVAICMLFVAINVKLISQPPQSQSVYGLFHSIDTADSGITLILIGLITLISGLVFCYQMKFYKAIVYLVANSIGILLVYVGFHRIAEASQIPSVPTFFLILQGLGIGTCLHWKSHRRLHQFLYMLAGLTAI FILNLPDWTVMALTAISFWDIVALTPCGLKMLVETANRRGDHDFPALLYNSSSY VNEVSDPTTNSGTPLTEFNSSSSRLSDSLRPVPIRQIREVREVGTRLGM GDFEYSLMGLNTQTCTPLTPVACFVSNLVGLITLITPLIVLSQTALPALFPFLAIAA IFYSSHALTPFDLCTSQILILI"  
 BASE COUNT 350 a 242 c 214 g 376 t  
 ORIGIN

Query Match 1.5%; Score 21; DB 34; Length 1182;  
 Best Local Similarity 100.08; Pred. No. 1.3; Mismatches 0; Indels 0; Gaps 0;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1378 tacattgaaaaa 1398  
 |||||  
 Db 1148 TACATTGAAAAA 1168

RESULT 8  
 RPAJ2554  
 LOCUS RPAJ2554 1267 bp mRNA VRT 12-AUG-1999  
 DEFINITION Rana perez1 mRNA for class IV-like alcohol dehydrogenase.  
 ACCESSION AJ002554  
 VERSION AJ002554.1 GI:2695886  
 KEYWORDS ADH gene; alcohol dehydrogenase.  
 SOURCE Perez's frog.  
 ORGANISM Rana perez1  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.  
 REFERENCE 1 (bases 1 to 1267)  
 AUTHORS Peralba,J.M., Cederlund,E., Crosas,B., Moreno,A., Julia,P., Martinez,S.E., Persson,B., Farres,J., Farres,X. and Joernvall,H.  
 TITLE Structural and enzymatic properties of a gastric NADP(H)-dependent alcohol dehydrogenase  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1267)  
 AUTHORS Crosas Navarro,B.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-NOV-1997) Crosas Navarro B., Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Facultat de Ciències, Dep. Bioquímica i Biologia Molecular, UAB., 08193, SPAIN

REMARK revised by [3]  
 REFERENCE 3 (bases 1 to 1267)  
 AUTHORS Crosas Navarro,B.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-DEC-1997) Crosas Navarro B., Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Facultat de Ciències, Dep. Bioquímica i Biologia Molecular, UAB., 08193, SPAIN  
 COMMENT On Dec 18, 1997 this sequence version replaced gi:2653282.  
 FEATURES  
 Location/Qualifiers  
 1. .1267  
 /organism="Rana perez1"  
 /db\_xref="taxon:8403"  
 /dev\_stage="adult"  
 /tissue\_type="stomach"  
 4. .1125  
 /gene="ADH"  
 4. .1125  
 /gene="ADH"  
 /EC\_number="1.1.1.1"  
 /codon\_start=1  
 /product="class IV-like alcohol dehydrogenase"  
 /protein\_id="CAA05553.1"  
 /db\_xref="GI:2695887"  
 /db\_xref="SPTREMBL:O57380"  
 /translation="MCTAGKDITCKAAVANEPHKLPLETITVAPPAHEVRKILAS GIGSDSVLKEIIPSPFPVILGHEAVGVESIGAGVTCVAPGDKVIPLFVPCGSCR ACKSSNFCENKMDGARTGLMADMTSRFTCRGKPIYNLVGTSTFTTYVVADIAVAR IDPKAPLESCLIGCGFATGYCAAVNTARVTPGSCAVFGLGGVGFSAIVGCAAGASR IIGVTHKDFPRAELGATECLNPKDYKPIYEVICETKNGVDYAVECAGRIETMM NALQSTVCGSGVTVLGLASPNRLDPLDLLTGRSLKSGVFGFGKEVSRVLVDY MKKINYNFLVSTKLTLQINKAPELLSSGGVRSIMY"  
 BASE COUNT 386 a 234 c 286 g 361 t  
 ORIGIN

Query Match 1.5%; Score 21; DB 4; Length 1267;  
 Best Local Similarity 100.08; Pred. No. 1.3; Mismatches 0; Indels 0; Gaps 0;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1378 tacattgaaaaa 1398  
 |||||  
 Db 1243 TACATTGAAAAA 1263

RESULT 9  
 MMU70324  
 LOCUS MMU70324 3258 bp mRNA ROD 02-OCT-1996  
 DEFINITION Mus musculus Fyn(T) (L-fyn(T)) mRNA, complete cds.  
 ACCESSION U70324  
 VERSION U70324.1 GI:1575676  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 3258)  
 AUTHORS Lee,C., Kim,M.-G., Jeon,S.H., Park,D.E., Park,S.D. and Seong,R.H.  
 TITLE Two Species of mRNAs for the fyn Proto-oncogene Are Produced by Alternative Polyadenylation in T cells  
 JOURNAL Unpublished (1996)  
 REFERENCE 2 (bases 1 to 3258)  
 AUTHORS Lee,C., Kim,M.-G., Jeon,S.H., Park,D.E., Park,S.D. and Seong,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-SEP-1996) IMBG, SNU, Shillim san 56-1, Kwanak, Seoul 151-742, S. Korea  
 FEATURES  
 Location/Qualifiers  
 1. .3258  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /cell\_type="T cell"  
 227. .1831  
 /gene="L-fyn(T)"  
 227. .1831

```
/gene="L-fyn(T)"
/function="protein tyrosine kinase"
/note="larger of two forms of fyn(T) produced by
alternative polyadenylation"
/codon_start=1
/product="Fyn(T)"
/protein_id="AAB09568.1"
/db_xref="GI:1575677"
/translation="MGCVCKDEAKLTERDGLNSOSSGYRYGTDTPTPOHPYFGV
TSIPNNFHAAGGGLTVFGGVNSSHTGTLTRGGTGVTILFVALYDYEARTEDDL
FKHKEFLINSSEGDWEARSLTGTGTYIPSNYVAPVDSIQAEWFYGLGRDAE
ROLLSGNFGFLRESQDGLNLTGAYSLSDIWDMDMKDHYKIRKLDNGGIYITR
AQCFETLQVHYSKAGCLCNLTWSSCTPQTSGLAKDAWEVARDLFLKXG
GCFEVLWGTNGNTKVAIKLPGTMSPESELEEAQIMKKLHDKLVQLYAVVSEEP
IIVTEYMSKSLDGLDKDEGRALKPLNVDMAQVAAGMAYIERMYIHRDLRSAN
ILVNGELICKIADFLARLIEDNEYTARGAKFFIKWTAPEAALYGRITIKSDVWSFG
ILLTELVLKGRVYPCGMNREVLEQVGRYRMPQDCPISLHLMHCKWKDPERP
TFEYLGQFLEDTFATPEQYQPGENL"
BASE COUNT      904 a  731 c  782 g  841 t
ORIGIN
```

```
Query Match      1.5%; Score 21; DB 12; Length 3258;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1378 tacattgaaaaa 1398
|||||
DB 3231 TACATTGAAAAA 3251
```

```
RESULT 10
AB025259
LOCUS      3323 bp mRNA ROD 28-SEP-1999
DEFINITION Mus musculus mRNA for granuphilin-b, complete cds.
ACCESSION AB025259
VERSION AB025259.1 GI:5926737
KEYWORDS granuphilin-b.
SOURCE Mus musculus pancreatic beta cell cell_line:MIN6 or betaHC9 cDNA to mRNA.
```

```
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS      Wang,J., Takeuchi,T., Yokota,H. and Izumi,T.
TITLE         Novel rabphilin-3-like protein associates with insulin-containing
              granules in pancreatic beta cells
JOURNAL       J. Biol. Chem. 274 (40), 28542-28548 (1999)
MEDLINE
REFERENCE     99428529
              2 (bases 1 to 3323)
AUTHORS       Izumi,T.
TITLE         Direct Submission
JOURNAL       Submitted (23-MAR-1999) to the DDBJ/EMBL/GenBank databases. Tetsuro
              Izumi, Institute for Molecular and Cellular Regulation, Gunma
              Univ., Department of Molecular Medicine; 3-39-15 Showa-machi,
              Maebashi, Gunma 371-8512, Japan
              (E-mail:tizumi@akagi.sb.gunma-u.ac.jp, Tel:+81-27-220-8856,
              Fax:+81-27-220-8896)
              Location/Qualifiers
                1. .3323
                  /organism="Mus musculus"
                  /db_xref="taxon:10090"
                  /cell_line="MIN6 or betaHC9"
                  /cell_type="pancreatic beta cell"
                  186. .1694
                  /codon_start=1
                  /product="granuphilin-b"
                  /protein_id="BAA84657.1"
                  /db_xref="GI:5926738"
                  /translation="MSEILDLSFLSEMERDLILGVLQORDEELRKADKRIIRLKNELL
                  EIKRKAGRSQHSYDRFCARQOEGRLIPKSTCGVGNHLVCRECVLESNGSWRC
                  KVCSEIELKATGDFYDQKVRNFDYRTGSEIRMSLRQPAVNKRRTAGQSLLOQT
                  QMGDIWPGRIIEQOQREQSVLFVFPKTRSGKSALEAESLDSYTDSDTSRSDS
```

```
FEATURES
source
CDS
Query Match      1.5%; Score 21; DB 12; Length 3741;
BASE COUNT      1033 a  778 c  946 g  984 t
ORIGIN
```

```
LDKSLPPEKKMSAPKSQVEKEIPPNQNAVCGDEGMFKKNTKKVLRSEYTKSV
IDLRPEVAQESGILGRSKSVPLGSLVDMEEEDIDHLVKLHROKLARGSMQSG
SSMTLGSIMSIYSEAGDFGNISVTGKIATFSLKFEQTQTILVIHVKECHQLAYADEAK
KSNPNYKTYLLPDKSRQGRKTSIKRDTINPLYDETFRYEISLSLAQRTLQFSVWH
HGRFRNTEFLGEAEVHMDSMKDKLDHCLPLHCKGSMAKWGTWIRLVKK"
BASE COUNT      918 a  680 c  826 g  899 t
ORIGIN
```

```
Query Match      1.5%; Score 21; DB 12; Length 3323;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1378 tacattgaaaaa 1398
|||||
DB 3298 TACATTGAAAAA 3318
```

```
RESULT 11
AB025258
LOCUS      3741 bp mRNA ROD 28-SEP-1999
DEFINITION Mus musculus mRNA for granuphilin-a, complete cds.
ACCESSION AB025258
VERSION AB025258.1 GI:5926735
KEYWORDS granuphilin-a.
SOURCE Mus musculus pancreatic beta cell cell_line:MIN6 or betaHC9 cDNA to mRNA.
```

```
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS      Wang,J., Takeuchi,T., Yokota,H. and Izumi,T.
TITLE         Novel rabphilin-3-like protein associates with insulin-containing
              granules in pancreatic beta cells
JOURNAL       J. Biol. Chem. 274 (40), 28542-28548 (1999)
MEDLINE
REFERENCE     99428529
              2 (bases 1 to 3741)
AUTHORS       Izumi,T.
TITLE         Direct Submission
JOURNAL       Submitted (23-MAR-1999) to the DDBJ/EMBL/GenBank databases. Tetsuro
              Izumi, Institute for Molecular and Cellular Regulation, Gunma
              Univ., Department of Molecular Medicine; 3-39-15 Showa-machi,
              Maebashi, Gunma 371-8512, Japan
              (E-mail:tizumi@akagi.sb.gunma-u.ac.jp, Tel:+81-27-220-8856,
              Fax:+81-27-220-8896)
              Location/Qualifiers
                1. .3741
                  /organism="Mus musculus"
                  /db_xref="taxon:10090"
                  /cell_line="MIN6 or betaHC9"
                  /cell_type="pancreatic beta cell"
                  186. .2207
                  /codon_start=1
                  /product="granuphilin-a"
                  /protein_id="BAA84656.1"
                  /db_xref="GI:5926736"
                  /translation="MSEILDLSFLSEMERDLILGVLQORDEELRKADKRIIRLKNELL
                  EIKRKAGRSQHSYDRFCARQOEGRLIPKSTCGVGNHLVCRECVLESNGSWRC
                  KVCSEIELKATGDFYDQKVRNFDYRTGSEIRMSLRQPAVNKRRTAGQSLLOQT
                  QMGDIWPGRIIEQOQREQSVLFVFPKTRSGKSALEAESLDSYTDSDTSRSDS
                  IDLRPEVAQESGILGRSKSVPLGSLVDMEEEDIDHLVKLHROKLARGSMQSG
                  SSMTLGSIMSIYSEAGDFGNISVTGKIATFSLKFEQTQTILVIHVKECHQLAYADEAK
                  KSNPNYKTYLLPDKSRQGRKTSIKRDTINPLYDETFRYEISLSLAQRTLQFSVWH
                  HGRFRNTEFLGEAEVHMDSMKDKLDHCLPLHCKGSMAKWGTWIRLVKK"
                  PASKLPVGGDRKSKSGGEGLOVMKAKNLTAAKSGTSDSFYKGLLPMRNKASK
                  RKTVPWKTLSPHYNHTVYNGVRLELDQHCLELTVMWDRPLASNDFLGGVRLGVTG
                  GISGEYVDWDSGEEVSLWQKMRQYFGSNAEGTLQLRSSNVKQKLGV"
BASE COUNT      1033 a  778 c  946 g  984 t
ORIGIN
```

Query Match 1.5%; Score 21; DB 12; Length 3741;

```

Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1378 tacattgaaaaaaaaaaaaa 1398
|||||
Db 3716 TACATTGAAAAAAAAAAAAA 3736

RESULT 12
AC005381 41378 bp DNA PRI 04-AUG-1998
DEFINITION Homo sapiens chromosome 19, cosmid F19926, complete sequence.
ACCESSION AC005381
VERSION AC005381.1 GI:3386588
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 41378)
AUTHORS Lamerdin,J.E., McCready,P.M., Skowronski,E., Adamson,A.W.,
Burkhardt-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stillwagen,S.,
Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J.,
Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
Kronmiller,B., Arellano,A., Montgomey,M., Ow,D., Nolan,M.,
Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
TITLE Sequence analysis of an ~600 kb contig from human 19q12 between
D19S766 and D19S326
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 41378)
AUTHORS Lamerdin,J.E.
TITLE Direct Submission
JOURNAL Submitted (04-AUG-1998) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT Map and sequence oriented from q centromere to telomere. Cosmid
F19926 is separated from cosmid R26169 to the left by an
approximately 3- 5 kb gap and from cosmid R31372 to the right by <1
kb. Additional map and sequence information may be obtained at:
http://www-bio.lnl.gov/dbbrp/genome/genome.html.
FEATURES
Location/Qualifiers
1..41378
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="F19926"
/chromosome="19"
/map="19q12 between D19S766 and D19S326"
/cell_line="UV5HL9-5B"
/clone_lib="LL19NC02 F chromosome 19-specific cosmid
library"
/note="Cosmid library constructed at LNL from flow-sorted
chromosomes from human-hamster hybrid UV5HL9-5B, which
carries chromosome 19 as its only human chromosome."
1859..2000
/note="Predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 78.000"
complement(2012..2113)
/rpt_family="MER5B"
2286..2586
/rpt_family="AluSp"
repeat_region 3188..3872
/rpt_family="MER51B"
4014..4114
/note="Predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 88.000"
complement(4129..4205)
/rpt_family="(GAAA)n"
4835..4927
/rpt_family="LTR9"
repeat_region 5117..5428
/rpt_family="LTR9"
repeat_region 5558..5645
/rpt_family="LTR9"

/rpt_family="MER65_internal"
5827..5946
/rpt_family="MER4_internal"
5983..6072
/rpt_family="MER4A2"
6073..6223
/rpt_family="MER4A"
complement(6570..6602)
/rpt_family="At_rich"
complement(7370..7671)
/rpt_family="AluSc"
7856..8027
/rpt_family="MER57_internal"
8009..8475
/rpt_family="MER57_internal"
8514..10630
/rpt_family="MER57_internal"
complement(10689..10971)
/rpt_family="AluSx"
10977..11122
/note="DSS similarity to AA523922 ng24g02.s1 NCI_CGAP_Co3
Homo sapiens cDNA clone IMAGE:935762; (564..421); 91%
identity."
11254..11890
/rpt_family="MER51B"
11400..11809
/note="DSS similarity to overlapping ESTs:-(11400..11809)
AA523922 ng24g02.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone
IMAGE:935762; (420..12); 99% identity --(11813..11512)
AA502625 ne42c08.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone
IMAGE:900014; Score: 573 identity: 301/302 (99%)."
11929..11996
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 55.000"
complement(12194..12429)
/rpt_family="MSTB"
complement(12463..12549)
/rpt_family="LIPAL16"
complement(12569..12746)
/rpt_family="MSTB"
complement(14253..14360)
/rpt_family="MIR"
14956..15266
/rpt_family="AluJo"
16003..16059
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 52.000"
complement(16203..16289)
/rpt_family="(GAAA)n"
complement(16290..16455)
/rpt_family="AluSg/x"
16458..16521
/rpt_family="AluJ/FRAM"
complement(17096..17619)
/rpt_family="MER34"
17886..18179
/rpt_family="AluJb"
18196..18358
/rpt_family="(GAAA)n"
complement(20161..20222)
/rpt_family="LINE2"
20660..20935
/rpt_family="AluSg"
complement(21051..21140)
/rpt_family="MER80"
complement(21152..21429)
/rpt_family="AluY"
complement(21441..21827)
/rpt_family="MER80"
complement(22215..22334)
/rpt_family="(GA)n"
23364..23468
/note="predicted exon, program: grail2exons_human_1.3,

```

```

frame: 0, quality: excellent, score: 100.000"
/repeat_region complement(23528..23599)
/repeat_region /rpt_family="(TA)n"
/repeat_region 24126..24195
/repeat_region /rpt_family="L1MA9"
/repeat_region 24214..24333
/repeat_region /rpt_family="(GAA)n"
/repeat_region 24355..24681
/repeat_region /rpt_family="L1MA9"
/repeat_region complement(26010..26037)
/repeat_region 26046..26347
/repeat_region /rpt_family="AT_rich"
/repeat_region /rpt_family="AluX"
/repeat_region 26796..26874
/repeat_region /rpt_family="(GGA)n"
/repeat_region 27062..27116
/repeat_region /rpt_family="(GAA)n"
/repeat_region complement(28049..28227)
/repeat_region /rpt_family="MIR"
/repeat_region complement(28300..28467)
/repeat_region /rpt_family="MIR"
/repeat_region complement(28969..29208)
/repeat_region /rpt_family="MER21B"
/repeat_region 29212..29505
/repeat_region /rpt_family="AluJo"
/repeat_region complement(29512..29936)
/repeat_region /rpt_family="MER21B"
/repeat_region 30362..30526
/repeat_region /rpt_family="L1MB3"
/repeat_region complement(31089..31183)
/repeat_region /rpt_family="MERSA"
/repeat_region 31973..32354
/repeat_region /rpt_family="THEIC"
/repeat_region complement(32727..32771)
/repeat_region /rpt_family="LINE2"
/repeat_region complement(32791..33051)
/repeat_region /rpt_family="AluY"
/repeat_region 33069..33278
/repeat_region /rpt_family="AluJb"
/repeat_region 33941..34302
/repeat_region /rpt_family="MLT1A1"
/repeat_region 34496..34662
/repeat_region /rpt_family="LINE2"
/repeat_region complement(34764..34869)
/repeat_region /rpt_family="LINE2"
/repeat_region complement(35765..36071)
/repeat_region /rpt_family="AluJo"
/repeat_region complement(36142..36226)
/repeat_region /rpt_family="L1M4"
/repeat_region 36586..36652
/repeat_region /note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 94.000"
/repeat_region complement(37147..37309)
/repeat_region /rpt_family="MERSB"
/repeat_region 37690..38218
/repeat_region /rpt_family="L1M4"
/repeat_region complement(38290..38313)
/repeat_region /rpt_family="AT_rich"
/repeat_region 38318..38618
/repeat_region /rpt_family="AluY"
/repeat_region complement(38830..39127)
/repeat_region /rpt_family="AluSq"
/repeat_region complement(39132..39418)
/repeat_region /rpt_family="AluJb"

Query Match 1.5%; Score 21; DB 11; Length 41378;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 900 ccaaggccacacagctgagtc 920
|||||
Db 40309 CCAAGGCCACACAGCTGAGTC 40329

```

```

RESULT 13
AC016692
LOCUS
DEFINITION
Homo sapiens clone RP11-93E22, WORKING DRAFT SEQUENCE, 1 unordered
pieces.
AC016692
AC016692.1 GI:6524386
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 160969)
AUTHORS
Waterston,R.H.
TITLE
The sequence of Homo sapiens clone
JOURNAL
Unpublished
2 (bases 1 to 160969)
AUTHORS
Waterston,R.H.
TITLE
Direct Submission
JOURNAL
Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
Center Project name: H_NH0093E22
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 160969: contig of 160969 bp in length.
FEATURES
Location/Qualifiers
1..160969
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-93E22"
BASE COUNT 43574 a 35321 c 36164 g 45748 t 162 others
ORIGIN

```

```

Query Match 1.5%; Score 21; DB 44; Length 160969;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 900 ccaaggccacacagctgagtc 920
|||||
Db 82286 CCAAGGCCACACAGCTGAGTC 82306

RESULT 14
AC006265/c
LOCUS
DEFINITION
Homo sapiens chromosome 17, clone hRPX.566_B_16, complete sequence.
AC006265
AC006265.1 GI:4199962
VERSION
HTG.
KEYWORDS
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 177707)
AUTHORS
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE
Unpublished
2 (bases 1 to 177707)
AUTHORS
Birren,B., Linton,L., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Baker,J., Baldwin,J., Collangelo,M., Collins,S., Collymore,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,K., Devon,K., Dewar,K.,
Cooke,P., Dearellano,K., Depayre,E., Forrest,C., Funke,R.,
Doneelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A.,

```

repeat_region	12608. .12670
repeat_region	/rpt_family="AT_rich"
repeat_region	complement(12721. .12854)
repeat_region	/rpt_family="MER5B"
repeat_region	13193. .13311
repeat_region	/rpt_family="FLAW_C"
repeat_region	13412. .13608
repeat_region	/rpt_family="HALI"
repeat_region	13704. .13738
repeat_region	/rpt_family="AT_rich"
repeat_region	14011. .14042
repeat_region	/rpt_family="(A)n"
repeat_region	14658. .14678
repeat_region	/rpt_family="AT_rich"
repeat_region	16354. .16384
repeat_region	/rpt_family="(A)n"
repeat_region	16680. .16871
repeat_region	/rpt_family="L2"
repeat_region	complement(17255. .17559)
repeat_region	/rpt_family="Alu"
repeat_region	17974. .18069
repeat_region	/rpt_family="MER63"
repeat_region	18779. .18821
repeat_region	/rpt_family="(CA)n"
repeat_region	complement(19927. .19991)
repeat_region	/rpt_family="L2"
repeat_region	20695. .23425
repeat_region	/rpt_family="Tigger2"
repeat_region	24227. .24442
repeat_region	/rpt_family="L2"
unsure	complement(24792. .24793)
repeat_region	/note="Single-stranded coverage."
repeat_region	25730. .25751
repeat_region	/rpt_family="AT_rich"
repeat_region	26342. .26724
repeat_region	/rpt_family="MLT2CA"
repeat_region	26809. .26922
repeat_region	/rpt_family="HERVL"
repeat_region	26923. .27210
repeat_region	/rpt_family="Alu"
repeat_region	27211. .29221
repeat_region	/rpt_family="HERVL"
repeat_region	29222. .29242
repeat_region	/rpt_family="(CAAA)n"
repeat_region	29243. .30893
repeat_region	/rpt_family="HERVL"
repeat_region	30905. .30925
repeat_region	/rpt_family="(TGGGGG)n"
repeat_region	30986. .31023
repeat_region	/rpt_family="AT_rich"
repeat_region	31088. .31131
repeat_region	/rpt_family="AT_rich"
repeat_region	complement(31212. .31307)
repeat_region	/rpt_family="MER5A"
repeat_region	31235. .31326
repeat_region	/rpt_family="MER5A"
repeat_region	31327. .31614
repeat_region	/rpt_family="AluSk"
repeat_region	31615. .31658
repeat_region	/rpt_family="MER5A"
repeat_region	complement(31669. .32037)
repeat_region	/rpt_family="LTR16A1"
repeat_region	32293. .32323
repeat_region	/rpt_family="AT_rich"
repeat_region	33503. .33824
repeat_region	/rpt_family="LTR16A1"
repeat_region	33834. .34062
repeat_region	/rpt_family="HERV16"
repeat_region	complement(34474. .34650)
repeat_region	/rpt_family="MIR"
repeat_region	complement(35534. .35711)
repeat_region	/rpt_family="MIR"
repeat_region	35816. .35942

```
FEATURES
source
Location/Qualifiers
1..177707
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HRPK.566_B_16"
/clone_lib="RPCI-11 human BAC library"
/map="17"
/chromosome="17"
1373..1423
/rpt_family="AT_rich"
repeat_region
2292..2323
/rpt_family="(TTTA)n"
repeat_region
complement(2747..2903)
/rpt_family="MIR"
repeat_region
3195..3237
/rpt_family="AT_rich"
repeat_region
3610..3643
/rpt_family="(TA)n"
repeat_region
3645..3679
/rpt_family="(TC)n"
repeat_region
6623..6681
/rpt_family="AT_rich"
repeat_region
7148..7175
/rpt_family="AT_rich"
repeat_region
7382..7432
/rpt_family="AT_rich"
repeat_region
7601..7654
/rpt_family="AT_rich"
repeat_region
7921..8055
/rpt_family="AT_rich"
repeat_region
8671..8809
/rpt_family="MERSB"
repeat_region
11018..11357
/rpt_family="LIM4"
unsure
11170..11180
/note="Five subclones have 11 A's here; one subclone has
12 A's here."
12112..12144
/rpt_family="AT_rich"
repeat_region
```

```
repeat_region      /rpt_family="MER94"
repeat_region      36401..36471
repeat_region      /rpt_family="MIR"
repeat_region      complement(36709..36946)
repeat_region      /rpt_family="LTR168"
repeat_region      complement(38848..40103)
repeat_region      /rpt_family="HSMAR1"
repeat_region      40104..40196
repeat_region      /rpt_family="(TA)n"
repeat_region      40771..41157
repeat_region      /rpt_family="MSTA"
repeat_region      41375..41516
repeat_region      /rpt_family="(TA)n"
repeat_region      41521..41577
repeat_region      /rpt_family="(CATATA)n"
repeat_region      41580..41613
repeat_region      /rpt_family="(CA)n"
repeat_region      42286..42313
repeat_region      /rpt_family="AT_rich"
repeat_region      42568..42647
repeat_region      /rpt_family="L2"
repeat_region      43518..43601
repeat_region      /rpt_family="AT_rich"
repeat_region      complement(43960..47295)
repeat_region      /rpt_family="LIME1"
repeat_region      47578..48195
repeat_region      /rpt_family="LIMEC"
repeat_region      48251..49317
repeat_region      /rpt_family="LIMEC"
repeat_region      49321..49649
repeat_region      /rpt_family="LIME2"
repeat_region      49804..49833
repeat_region      /rpt_family="AT_rich"
repeat_region      complement(50350..50523)

Query Match      1.5%; Score 21; DB 40; Length 177707;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1378 tacatttgaaaaa 1398
|||||
Db 26336 TACATTGAAAAA 26316
```

```
RESULT 15
AC006427 AC006427 178524 bp DNA PRI 16-NOV-1999
LOCUS Homo sapiens chromosome 4 clone C0201M04, complete sequence.
DEFINITION AC006427
ACCESSION AC006427
VERSION AC006427.13 GI:6437518
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 178524)
AUTHORS Stone,N.E., Schmutz,J.J., Shang,J., Pennacchio,L.A., Cox,D.R. and
Myers,R.M.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 178524)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1999) Department of Genetics, Stanford Human
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
REFERENCE 3 (bases 1 to 178524)
AUTHORS Stone,N.E., Schmutz,J.J., Shang,J., Pennacchio,L.A., Cox,D.R. and
Myers,R.M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Department of Genetics, Stanford Human
Genome Center, 4005 Miranda Avenue, Palo Alto, CA 94304, USA
COMMENT On Nov 16, 1999 this sequence version replaced gi:5861531.
Quality: Phrap Quality >=40 99.9% of sequence;
```

```
Estimated Total Number of Errors is 0.1.
STS Content:
SHGC-33956 G29584
SHGC4-1650 G05155
FEATURES
source
1..178524
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone_lib="ROSWELL PARK CANCER RPCI - 11 Human Male BAC
Library"
BASE COUNT 55602 a 36515 c 35601 g 50806 t
ORIGIN
```

Query Match 1.5%; Score 21; DB 40; Length 178524;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1378 tacatttgaaaaa 1398  
|||||  
Db 109166 TACATTGAAAAA 109186

Search completed: May 1, 2000, 16:16:43  
Job time: 19103 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 18:50:48 ; Search time 250.34 Seconds  
(without alignments)  
1397.174 Million cell updates/sec

Title: US-09-215-435-71

Perfect score: 1398  
Sequence: 1 gatcgctgagggcccgagc.....acatttgaaaaa 1398

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 311585 seqs, 125096042 residues

Word size : 0

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	5.2	203	1	X41456 Human secreted pro
2	20	1.4	863	1	Q06403 Sequence encoding
3	20	1.4	863	1	Q28683 PI-linked LFA-3 us
4	20	1.4	1009	1	Q06404 Plasmid T4/LFA-3/2
5	20	1.4	1078	1	Q06405 Plasmid T4/LFA-3/2
6	19	1.4	1101	1	X23750 Rice antranilate
c 7	19	1.4	1268	1	T74891 Human neurogenic d
c 8	19	1.4	1268	1	V42932 DNA encoding human
9	19	1.4	1498	1	X23749 Rice antranilate
10	19	1.4	1970	1	Q97805 Clone pygama/ cod
11	19	1.4	2245	1	T84445 Human STCH chapero
12	18	1.3	115	1	T39461 Growth regulatory
13	18	1.3	115	1	T39461 Human growth regul
14	18	1.3	199	1	Q42772 Ligand-induced gen
15	18	1.3	428	1	V34202 Human secreted pro
16	18	1.3	594	1	T16870 Pepper Group 2 pro
17	18	1.3	603	1	V73486 Human SRE-ZBP anal
18	18	1.3	880	1	T91714 Rat Smlm/CRP2 cDN
19	18	1.3	1023	1	Q28271 Encodes human IGF
20	18	1.3	1023	1	T15231 Insulin-like growt
21	18	1.3	1023	1	T94634 TNF-RI-DD ligand p
c 22	18	1.3	1025	1	X30319 DNA encoding a hum
23	18	1.3	1045	1	V54583 Human secretory pr
24	18	1.3	1051	1	V73487 Human SRE-ZBP anal
25	18	1.3	1065	1	X30407 DNA encoding a hum
c 26	18	1.3	1180	1	X14017 H. pylori GHPD 75
27	18	1.3	1233	1	T16869 Pepper Group 2 con
28	18	1.3	1683	1	V73484 Human SRE-ZBP anal
29	18	1.3	1727	1	T16871 Pepper Group 2 gen
30	18	1.3	1858	1	X27169 Rat GMEB-2' coding
31	18	1.3	1860	1	T31929 Retinoid x recepto
32	18	1.3	2098	1	X02558 Human B1 cDNA. New
33	18	1.3	2168	1	V73485 Human SRE-ZBP anal

#### ALIGNMENTS

##### RESULT 1

X41456  
ID AC X41456 standard; cDNA; 203 BP.  
AC X41456;  
DE 22-JUN-1999 (first entry)  
DE Human secreted protein 5', EST SEQ ID NO: 115 from WO 9906553.  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; antiinflammatory; tumour inhibition; antitumour; ds.  
OS Homo sapiens.  
PN WO9906553-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; IB1237.  
PI 01-AUG-1997; US-905051.  
PA (GSET ) GENSET.  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
DR WPI: 99-153783/13.  
DR P-PSDB: Y12598.  
PT New nucleic acids encoding human secreted proteins - obtained from  
PT cDNA libraries derived from umbilical cord, lymph ganglia,  
PT lymphocytes and placental tissue  
PS Claim 1: Page 225; 411pp; English.  
CC X41379 to X41526 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins, and encode the proteins given in Y12521 to Y12668,  
CC respectively. The proteins given represent the signal peptide and an  
CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
CC can be used for producing secreted human gene products. They can also  
CC be used to develop products for diagnosis and therapy. The proteins  
CC obtained may have cytokine activity, cell proliferation/differentiation  
CC activity, haematopoiesis regulating activity, tissue growth regulating  
CC activity, reproductive hormone regulating activity, chemotactic/  
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
CC ligand activity, antiinflammatory activity, tumour inhibition activity  
CC or other activities. The products can be used in forensic, gene therapy  
CC and chromosome mapping procedures. The sequences can also be used for  
CC obtaining corresponding promoter sequences. The nucleic acids encoding  
CC the signal peptide can be used for directing extracellular secretion of  
CC a polypeptide or the insertion of a polypeptide into a membrane, or  
CC importing a polypeptide into a cell.  
SQ Sequence 203 BP; 32 A; 68 C; 66 G; 36 T;

Query Match 5.2%; Score 72; DB 1; Length 203;  
Best Local Similarity 100.0%; Pred. No. 6.9e-23;  
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 tatgaatgagcctcagggcgctgttgcggcgctgttcggagagtcctctgtcag 99

DB 115 TGTGAATGAGGCTGCAGGGCGCTGTGTTCGGCGCTCGGTGAGAGTCTCTGTCTGAG 174

QY 100 ccgaacacctga 111

DB 175 CCGCAACCTCTGA 186

34 18 1.3 2172 1 V16852 Nicotiana glutinos  
35 18 1.3 2452 1 Q30966 Encodes vitamin K  
36 18 1.3 3138 1 Q11712 Shuttle vector pMU  
37 18 1.3 3157 1 Q13316 Duffey receptor gen  
38 18 1.3 3171 1 V32920 Solanum tuberosum  
39 18 1.3 3525 1 Q27658 N-sam cDNA. Gene p  
40 18 1.3 3539 1 V11783 H. contortus pHC-1  
41 18 1.3 4053 1 X52266 Protein PRO326 cDN  
42 18 1.3 4177 1 V19980 Full length Kerat1  
c 43 18 1.3 10282 1 V09023 Homo sapiens 20q13  
c 44 18 1.3 235033 1 V57926 Hereditary haemoch  
c 45 18 1.3 237326 1 V57903 Hereditary haemoch

```

RESULT 2
Q06403 Q06403 standard; DNA; 863 BP.
AC Q06403;
DT 06-FEB-1991 (first entry)
DE Sequence encoding lymphocyte function-associated antigen (LFA-3)
DE with phosphatidylinositol (PI) linkage signalling sequence.
KW Plasma membrane binding affinity; micelle; ds.
FT Key Location/Qualifiers
FT cds 18..737
FT /*tag= a
FT 585..737
FT /*tag= b
FT misc_signal
FT /label= Phosphatidylinositol linkage signal sequence
PN WO9012099-A.
PD 18-OCT-1990.
PF 05-APR-1990; U01859.
PR 10-APR-1989; US-335688.
PA (BIOG-) BIOGEN INC.
PI Wallner BP;
DR WPI; 90-334849/44.
DR P-PSDB; R07604.
PT from lymphocyte function-associated antigen 3, used for prodn. of
PT phosphatidyl-inositol linkage signalling DNA sequence - derived
PT chimeric proteins
PT Disclosure; Fig 2; 53pp; English.
CC The signal sequence is attached downstream of the LFA-3 sequence
CC encoding a secreted protein, which will then produce proteins
CC covalently anchored to the cell surface in which they were produced.
CC This can give rise to plasma membrane binding, enhanced
CC purifiability, micelle formation etc. especially useful in the
CC production of chimeric targeted drugs, to produce micellular or
CC liposomal delivery systems or in enhanced purification and screening
CC of cells, proteins or DNA libraries.
SQ Sequence 863 BP; 289 A; 161 C; 166 G; 247 T;

Query Match 1.4%; Score 20; DB 1; Length 863;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1379 acatttgaaaaa 1398
Db 822 ACATTGAAAAA 841

RESULT 4
Q06404 Q06404 standard; DNA; 1009 BP.
AC Q06404;
DT 06-FEB-1991 (first entry)
DE Plasmid T4/LFA-3/AD sequence with C-terminal encoding
DE Phosphatidylinositol (PI) linkage signalling sequence.
KW Plasma membrane binding affinity; micelle; ds.
FT Key Location/Qualifiers
FT cds 1..885
FT /*tag= a
FT misc_signal 799..885
FT /*tag= b
FT /label= Phosphatidylinositol linkage signal sequence
PN WO9012099-A.
PD 18-OCT-1990.
PF 05-APR-1990; U01859.
PR 10-APR-1989; US-335688.
PA (BIOG-) BIOGEN INC.
PI Wallner BP;
DR WPI; 90-334849/44.
DR P-PSDB; R07604.
PT Phosphatidyl-inositol linkage signalling DNA sequence - derived
PT from lymphocyte function-associated antigen 3, used for prodn. of
PT chimeric proteins
PT Disclosure; Fig 4; 53pp; English.
CC The signal sequence is attached downstream of the plasmid sequence
CC encoding a secreted protein, which will then produce proteins
CC covalently anchored to the cell surface in which they were produced.
CC This can give rise to plasma membrane binding, enhanced
CC purifiability, micelle formation etc. especially useful in the
CC production of chimeric targeted drugs, to produce micellular or
CC liposomal delivery systems or in enhanced purification and screening
CC of cells, proteins or DNA libraries.
SQ Sequence 1009 BP; 316 A; 215 C; 253 G; 225 T;

Query Match 1.4%; Score 20; DB 1; Length 1009;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1379 acatttgaaaaa 1398
Db 822 ACATTGAAAAA 841

RESULT 3
Q28683 Q28683 standard; DNA; 863 BP.
AC Q28683;
DT 17-FEB-1993 (first entry)
DE PI-linked LFA-3 used to make PIM3 deletion mutant.
DE Phosphatidylinositol anchored lymphocyte associated antigen-3;
KW T-lymphocyte accessory molecule; CD2 binding site; immunomodulator;
KW ss.
OS Homo sapiens.
FT Key Location/Qualifiers
FT misc_feature 369..581
FT /*tag= a
FT /label= PIM3_region
FT /note= "region deleted in mutant"
FT signal_peptide 18..101
FT /*tag= b
FT mat_peptide 102..737
FT /*tag= c
FT /product= PI-linked_LFA-3
PN EP-503648-A.
PD 16-SEP-1992.
PF 12-MAR-1992; 104320.
PR 12-MAR-1991; US-667971.
PR 07-OCT-1991; US-770967.

```

Db 970 ACATTGAAAAA 989

## RESULT 5

Q06405 Q06405 standard; DNA; 1078 BP.  
AC Q06405;  
DT 06-FEB-1991 (first entry)  
DE Plasmid T4/LFA-3/2 sequence with C-terminal encoding  
DE phosphatidylinositol (PI) linkage signaling sequence.  
KW Plasma membrane binding affinity; micelle; ds.  
FH Key  
FT cds  
FT misc\_signal  
FT 1..954  
FT 802..954  
FT /\*tag= a  
FT /\*tag= b  
FT /label= Phosphatidylinositol linkage signal sequence  
PN W09012099-A.  
PD 18-OCT-1990.  
PF 05-APR-1990; U01859.  
PR 10-APR-1989; US-335688.  
PA (BIOG-) BIOGEN INC.  
PI Wallner BP;  
DR WPI; 90-334849/44.  
DR P-PSDB; R07606.  
PT Phosphatidyl-inositol linkage signalling DNA sequence - derived  
PT from lymphocyte function-associated antigen 3, used for prodn. of  
PT chimeric proteins  
PS Disclosure; Fig 5; 53pp; English.  
CC The signal sequence is attached downstream of the Plasmid sequence  
CC encoding a secreted protein, which will then produce proteins  
CC covalently anchored to the cell surface in which they were produced.  
CC This can give rise to plasma membrane binding, enhanced  
CC purification, micelle formation etc. especially useful in the  
CC production of chimeric targeted drugs, to produce micellar or  
CC liposomal delivery systems or in enhanced purification and screening  
CC of cells, proteins or DNA libraries.  
SQ Sequence 1078 BP; 339 A; 232 C; 260 G; 247 T;

Query Match 1.4%; Score 20; DB 1; Length 1078;

Best Local Similarity 100.0%; Pred. No. 7.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1379 acatttgaaaaa 1398

Db 1039 ACATTGAAAAA 1058

## RESULT 6

X23750 X23750 standard; DNA; 1101 BP.  
AC X23750;  
DT 25-JUN-1999 (first entry)  
DE Rice antranilate synthase alpha-subunit promoter DNA.  
DE Antranilate synthase; alpha-subunit; ASA; rice; isozyme; plant; seed;  
KW maize; wheat; tryptophan content; nutritional value; promoter; ss.  
OS Oryza sativa.  
PN W0911800-A1.  
PD 11-MAR-1999.  
PF 31-AUG-1998; J03883.  
PR 29-AUG-1997; JP-235049.  
PA (HOKK) HOKKO CHEM IND CO LTD.  
PA (NORQ) JAPAN MIN AGRIC FORESTRY & FISHERIES.  
PI Hasegawa H, Terakawa T, Tozawa Y, Wakasa K;  
DR WPI; 99-228982/19.  
PT DNA encodes 'a'-subunit of first isozyme of rice antranilate  
PT synthase - used for improving tryptophan production and  
PT nutritional value of crops, e.g. rice, maize or wheat  
PS Example 5; Page 134-135; 152pp; Japanese.  
CC This invention describes a novel rice antranilate synthase first  
CC isozyme alpha-subunit. The encoding DNA can be used to produce  
CC transformant plants and seeds, of e.g. rice, maize or wheat, with

CC enhanced tryptophan content and nutritional value of the crops.  
SQ Sequence 1101 BP; 291 A; 239 C; 260 G; 311 T;

Query Match 1.4%; Score 19; DB 1; Length 1101;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1067 aagaagggaggaggaagg 1085

Db 98 AAGAAGGGAGGAGGAAGG 116

## RESULT 7

T74891/c T74891 standard; DNA; 1268 BP.  
AC T74891;  
DT 02-OCT-1997 (first entry)  
DE Human neurogenic differentiation protein (NeuroD3) DNA clone 20A1.  
DE Neurogenic differentiation protein; NeuroD; neuroD3 gene;  
KW transcriptional activator; neuron; pancreas; gastrointestinal;  
KW knock-out mouse; transgenic animal; cancer; diabetes; gene therapy;  
KW ss.  
OS Homo sapiens.  
FH Key  
FT cds  
FT Location/Qualifiers  
FT 55..768  
FT /\*tag= a  
PN W09716548-A1.  
PD 09-MAY-1997.  
PF 30-OCT-1996; U17532.  
PR 02-NOV-1995; US-552142.  
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
PA (WEIN/) WEINTRAUB N.  
PI Hollenberg SM, Lee JE, Tapscott SJ, Weintraub HM;  
DR WPI; 97-272117/24.  
DR P-PSDB; W22440.  
PT Nucleic acid encoding neurogenic differentiation polypeptide -  
PT useful e.g. in regulating neuronal, endocrine and gastrointestinal  
PT development  
PS Claim 1; Page 64-65; 81pp; English.  
CC Neurogenic differentiation (NeuroD) genes (T74887-94) and proteins  
CC (W22436-43) from human, mouse and frog have been identified,  
CC isolated and sequenced. NeuroD polypeptides are tissue-specific  
CC basic-helix-loop-helix (bHLH) transcriptional activators involved  
CC in neuronal, endocrine and gastrointestinal development. They were  
CC discovered by expression cloning and screening assays designed to  
CC identify possible bHLH proteins capable of interacting with the  
CC protein product of the Drosophila daughterless gene. Novel neuroD2  
CC and neuroD3 genes, related to neuroD1, have been identified.  
CC NeuroD nucleic acids can be used to produce NeuroD polypeptides, and  
CC construction of test cell lines, as probes, in gene therapy, and to  
CC produce transgenic animals as models of disease.  
SQ Sequence 1268 BP; 245 A; 455 C; 344 G; 224 T;

Query Match 1.4%; Score 19; DB 1; Length 1268;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 734 gggtcagagagcgaggagg 752

Db 650 GGGTCAGAGCGGGGAGG 632

## RESULT 8

V42932/c V42932 standard; DNA; 1268 BP.  
AC V42932;  
DT 21-OCT-1998 (first entry)  
DE DNA encoding human neuroD3 protein, which is a bHLH protein.  
DE Basic helix-loop-helix; bHLH; neuroD; neuroectodermal tumour;  
KW classification; medulloblastoma; human; ds.  
OS Homo sapiens.

```

FH Key Location/Qualifiers
FT CDS 55..768
FT /*tag= a
FT /product= neuroD3
PN US5795723-A.
PD 18-AUG-1998.
PF 07-AUG-1997; US-910973.
PR 07-AUG-1997; US-910973.
PR 06-MAY-1994; US-239238.
PR 02-NOV-1995; US-552142.
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PI Olson JM, Tapscott SJ.
DR WPI: 98-465661/40.
DR P-PSDB: W71016.
PT Classifying neuroectodermal tumours from expression pattern of
PT basic-helix-loop-helix genes - especially for identifying
PT medulla/blastoma and assessing its aggressiveness, specifically
PT associated with expression of bHLH genes neuroD 1-3
PS Example 11; Columns 61-64; 43pp; English.
CC The present sequence encodes a protein which is a member of the basic
CC helix-loop-helix (bHLH) protein family, and is designated neuroD3. The
CC bHLH genes are a family of genes associated with vertebrate neuronal,
CC endocrinal and gastrointestinal development. The observed pattern of
CC neuroD expression distinguishes subclasses of neuroectodermal tumours.
CC The specification describes a method for the classification of human
CC neuroectodermal tumours. The method comprises measuring, in a tumour
CC sample, expression of at least one basic bHLH gene and identifying the
CC tumour subclass by matching expression to predetermined expression
CC profiles for known subclasses. For classifying the tumour as a
CC medulloblastoma, the bHLH gene detected is neuroD1 and neuroD3.
CC The method is used to classify neuroectodermal tumours, and to identify
CC medulloblastoma and for prognosis of this as aggressive.
SQ Sequence 1268 BP; 245 A; 455 C; 344 G; 224 T;
Query Match 1.4%; Score 19; DB 1; Length 1268;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 734 gggtcagagcgaggagg 752
DB 650 GGCTCAGAGCGGGGAGG 632
RESULT 9
X23749
ID X23749 standard; DNA; 1498 BP.
AC X23749;
DT 25-JUN-1999 (first entry)
DE Rice antranilate synthase first isozyme alpha-subunit DNA fragment.
KW Anthranilate synthase; alpha-subunit; ASA; rice; isozyme; plant; seed;
KW maize; wheat; tryptophan content; nutritional value; ss.
OS Oryza sativa.
FH Key Location/Qualifiers
FT CDS 1102..1498
FT /*tag= a
FT /product= "ASA synthase"
FT /note= "Partial coding sequence. Incomplete exon 2"
FT exon 1102..1233
FT /*tag= b
FT /number= 1
FT intron 1234..1318
FT /*tag= c
FT /number= 1
FT exon 1318..1498
FT /*tag= d
FT /number= 2
FT /note= "partial exon 2 sequence"
PN W09911800-A1.
PD 11-MAR-1999.
PF 31-AUG-1998; J03883.
PR 29-AUG-1997; JP-235049.
PA (HOKK) HOKKO CHEM IND CO LTD.
PA (NORQ) JAPAN MIN AGRIC FORESTRY & FISHERIES.
PI Hasegawa H, Terakawa T, Tozawa Y, Wakasa K;
DR WPI: 99-228982/19.
DR P-PSDB: W93811.
PT DNA encodes ^a-subunit of first isozyme of rice antranilate
PT synthase - used for improving tryptophan production and
PT nutritional value of crops, e.g. rice, maize or wheat
PS Example 5; Page 131-133; 152pp; Japanese.
CC This invention describes a novel rice antranilate synthase first
CC isozyme alpha-subunit. The encoding DNA can be used to produce
CC transformant plants and seeds, of e.g. rice, maize or wheat, with
CC enhanced tryptophan content and nutritional value of the crops.
SQ Sequence 1498 BP; 343 A; 362 C; 404 G; 389 T;
Query Match 1.4%; Score 19; DB 1; Length 1498;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1067 aagaagggaggaggagg 1085
DB 98 AAGAAGGGGAGGAGGAGG 116
RESULT 10
Q97805
ID Q97805 standard; cDNA; 1970 BP.
AC Q97805;
DT 17-JUL-1996 (first entry)
DE Clone pGamma7 coding for mouse CCT-gamma subunit.
KW Chaperonin containing TCP-1; CCT; subunit; protein folding complex;
KW murine; denaturation; renaturation; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT CDS 102..1739
FT /*tag= a
FT /product= CCT-gamma
PN W09520654-A1.
PD 03-AUG-1995.
PF 31-JAN-1994; GB-001791.
PR 09-SEP-1994; GB-018234.
PA (CANC-) INST CANCER RES ROYAL CANCER HOSPITAL.
PI Ashworth A, Kubota H, Willison KR;
DR WPI: 95-275439/36.
DR P-PSDB: R79579.
PT Folding protein complex sub-unit(s) - with ability to form complex
PT in vitro, useful in facilitating folding of proteins
PS Claim 15; Fig 8c; 94pp; English.
CC Mouse cDNAs encoding novel TCP-1-related proteins were isolated
CC using a combination of methods. Clones pDelta2 and pZeta12 were
CC isolated by cross-hybridisation with C.elegans cDNA probes recovered
CC from a 5'-expressed sequence tag collection. The clones pBeta2,
CC pepsilon5 and ptheta1 were isolated by cross-hybridisation with
CC human probes recovered by PCR of human HT1080 cell line cDNA using
CC degenerate primers from a conserved region of TCP-1 and TF55. Clone
CC pCB80 was isolated during a mouse testis cDNA sequencing project
CC and clone pGamma7 was recovered by hybridisation with a mouse cDNA
CC PCR product which was made with primers derived from the sequence of
CC a TCP-1 related gene fragment recovered accidentally from a human
CC kidney cDNA library during a screen for ion transport channel
CC genes. TCP-1 and the proteins encoded by the cDNA clones are
CC individual subunits of a heteromeric chaperonin which has been
CC designated "chaperonin containing TCP-1" (CCT). The chaperonin acts
CC as a protein folding complex. The present sequence is that of cDNA
CC clone pGamma7 which encodes the CCT-gamma subunit.
SQ Sequence 1970 BP; 542 A; 445 C; 462 T;
Query Match 1.4%; Score 19; DB 1; Length 1970;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1380 catttgaaaaa 1398
DB 1952 CATTGAAAAA 1970

RESULT 11
T84445
ID T84445 standard; cDNA; 2245 BP.
AC T84445;
DE 17-NOV-1997 (first entry)
DE Human STCH chaperone protein cDNA.
KW STCH; stress 70 protein; chaperone protein; protein folding; human;
KW ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 37..1452
FT signal_peptide 37..102
FT mat_peptide 103..1449
FT  /*tag= a
FT  /*tag= b
FT  /*tag= c
FT US5646249-A.
PN 08-JUL-1997.
PD 08-JUL-1997.
PF 28-FEB-1994; 203905.
PR 28-FEB-1994; US-203905.
PA (US5646249) US DEPT HEALTH & HUMAN SERVICES.
PI Kaye FJ, Otterson GA;
DR WPI; 97-362996/33.
DR P-PSDB: W26311.
PT Recombinant Stress 70 chaperone family STCH proteins - useful as
PT chaperone proteins for facilitating protein folding
PS Example 1; Column 29-34; 30pp; English.
CC This cDNA sequence codes for a novel human stress 70 chaperone
CC family protein (W26311), designated STCH, that can be used to
CC facilitate protein folding. An initial STCH clone was obtd. by
CC screening a K562 erythroleukemia cell library with a probe based
CC on the retinoblastoma binding protein-1 gene, and further
CC overlapping clones were isolated from the K562 library and an H69
CC library using a partial STCH clone as probe. The STCH gene was
CC localised to chromosome 21q. It is induced by calcium ionophore
CC A23187 but not by heat shock. STCH nucleic acids can be used to
CC produce recombinant STCH polypeptides, to determine STCH copy
CC number in a cell genome, and in methods for identifying the long
CC arm of chromosome 21 for genetic analysis.
SQ Sequence 2245 BP; 711 A; 466 G; 669 T;

Query Match 1.48; Score 19; DB 1; Length 2245;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1380 catttgaaaaa 1398
DB 2221 CATTGAAAAA 2239

RESULT 12
T39461
ID T39461 standard; cDNA; 115 BP.
AC T39461;
DE 14-NOV-1996 (first entry)
DE Growth regulatory protein cDNA, GRP-24, isolated using Hel-N1.
KW Human; neuron-specific protein; Hel-N1; 3'-UTR instability sequence;
KW paraneoplastic sensory neuropathy; oncoprotein; lymphokine; rat; elav;
KW RNA recognition motif; RRM; Drosophila; cellular growth; localisation;
KW instability; translatability; neurons; autoimmune protein; PE; PCP; PSN;
KW central nervous system; cancer; paraneoplastic cerebellar degeneration;
KW paraneoplastic encephalomyelitis; RNP-1 octamer sequence; ss.
OS Synthetic.
PN US5525495-A.
PD 11-JUN-1996.
PF 11-JUN-1996.
PR 11-MAY-1992; 881075.
PR 11-MAY-1992; US-881075.

```

```

PR 15-SEP-1993; US-120827.
PA (YUDU-) UNIV DUKE.
PI Gao F, Keene JD, Levine T;
DR WPI; 96-286398/29.
PT Prod of cDNA library for related proteins - by screening total cell
PT mRNA with RNA-binding protein Hel-N1 or Hel-N2
PS Disclosure; Column 29-30; 66pp; English.
CC The sequences given in T39454-74 represent growth regulatory protein
CC cDNA's which were isolated due to their ability to bind to the human
CC neuron-specific protein, Hel-N1. The mRNA sequences corresponding to
CC these cDNA's all contained the characteristic short stretches of
CC uridylates found in the random RNA selection experiments (see also
CC T39400-52). These sequences are found near the poly A stretch and
CC within the 3' UTR. All the cDNA's isolated represented growth
CC regulatory proteins, including the ACK nonreceptor tyrosine kinase, the
CC src/yes type proto-oncogene and the B7G1 (P3/P4/interferon related)
CC growth regulatory protein. No house keeping or non-growth related genes
CC were detected. Hel-N1 was found to contain 3 RNA recognition motifs
CC (RRM's), where the third one (see also W00244) is sufficient for mRNA
CC 3'-UTR binding activity. Full length Hel-N1, when transfected into a
CC cell, causes cellular growth to cease, however, if just the third binding
CC domain is transfected into cells, the cells undergo rapid growth.
CC Hel-N1 binds as a multimer along the mRNA, presumably enhancing its
CC localisation, instability and/or regulating its translatability and/or
CC deadenylation it. This protein may be responsible for the growth
CC cessation of neurons. Hel-N1 is an autoimmune protein in certain
CC patients who show central nervous system manifestations of cancer
CC called paraneoplastic cerebellar degeneration (PCD), paraneoplastic
CC encephalomyelitis (PE) or paraneoplastic sensory neuropathy (PSN).
SQ Sequence 115 BP; 35 A; 26 C; 14 G; 40 T;

Query Match 1.38; Score 18; DB 1; Length 115;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1381 atttgaaaaa 1398
DB 97 ATTTGAAAAA 114

RESULT 13
V37479
ID V37479 standard; cDNA; 115 BP.
AC V37479;
DE 07-SEP-1998 (first entry)
DE Human growth regulatory protein (GRP) cDNA GRP-24.
KW Growth regulatory protein; Hel-N1; oncogene; cytokine; lymphokine;
KW chromosome mapping; human; functionally related protein; GRP; ss.
OS Homo sapiens.
PN US5773246-A.
PD 30-JUN-1998.
PF 07-JUN-1995; 478675.
PR 15-SEP-1993; US-120827.
PR 11-MAY-1992; US-881075.
PR 07-JUN-1995; US-478675.
PA (GAOF/) GAO F.
PA (KEEN/) KEENE J D.
PA (LEVI/) LEVINE T.
PI Gao F, Keene JD, Levine T;
DR WPI; 98-387003/33.
DE Use of proteins which bind RNA - for obtaining a cDNA library
DE containing members encoding structurally or functionally related
PT proteins from total cell mRNA.
PS Disclosure; Column 29; 67pp; English.
CC Sequences shown in V37472 to V37492 represent the growth regulatory
CC protein (GRP) cDNAs isolated from human brain mRNA using Hel-N1. The
CC invention provides a method for obtaining a cDNA library having members
CC encoding a group of structurally or functionally related proteins from
CC total cell mRNA. The method comprises binding RNA representing total cell
CC mRNA to a protein that binds RNA and has specific binding to untranslated
CC regions of a subset of the total cell mRNA, where the protein that binds
CC RNA has been purified to remove any other protein that binds RNA. The

```

CC resulting bound products are separated and a cDNA library is prepared.  
 CC The protein that binds RNA may be Hel-N1, Hel-N2, CARC, DT-7, K1, K2,  
 CC K3, Hub, hucP, rbp9, eIF4B, sxl, tra-2, AUBF, AUF, ASF/SF2, U2AF,  
 CC SC35, or hnRNP proteins. The cDNAs obtained can be used for chromosome  
 CC mapping and genome sequencing of the structurally or functionally related  
 CC genes encoding growth regulatory proteins, proto-oncogenes, cytokines,  
 CC lymphokines, or anti-oncogene proteins. They can also be used in  
 CC diagnostic methods for determining imbalances in such genes and the  
 CC efficacy of various treatments to correct any imbalance.  
 SQ Sequence 115 BP; 35 A; 26 C; 14 G; 40 T;

Query Match 1.3%; Score 18; DB 1; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1381 atttgagaaaaa 1398

Db 97 ATTTCAGAAAAA 114

#### RESULT 14

Q42772  
 ID Q42772 standard; cDNA; 199 BP.

AC Q42772; 1993 (first entry)

DT 24-SEP-1993 (first entry)

DE Ligand-induced gene, clone 10A8.

KW Lig; binding; library; receptor; growth phase transition;

KW clonal expansion; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT misc\_difference 92

FT /\*tag= a

FT /note= "residue represented as "N" in the

FT specification"

PN WO9310230-A.

PD 27-MAY-1993.

PF 18-NOV-1992; U10087.

PA 20-NOV-1991; US-796066.

PR (DART-) DARTMOUTH COLLEGE.

PI Smith KA;

DR WPI; 93-182547/22.

PT Preparing cDNA library enriched in ligand inducible genes - by

PT activating cells in presence of labelled RNA precursor and RNA

PT enhancer, isolating labelled RNA, converting to DNA and screening

PT for ligand induction

PS Claim 28; Page 36; 71pp; English.

CC A cDNA library of clones enriched in genes induced by ligand

CC binding were prep. Some of these clones contain the DNA

CC sequences given in Q42769-87.

CC An enriched library facilitates identification and characterisation

CC of ligand-inducible genes (LIG) that triggered immediately and/or

CC shortly after receptor activation. These genes may be involved in

CC stimulating growth phase transitions and subsequent clonal expansion.

SQ Sequence 199 BP; 84 A; 25 C; 26 G; 63 T;

Query Match 1.3%; Score 18; DB 1; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1381 atttgagaaaaa 1398

Db 180 ATTTCAGAAAAA 197

#### RESULT 15

V34202

ID V34202 standard; DNA; 428 BP.

AC V34202;

DT 28-JAN-1999 (first entry)

DE Human secreted protein gene 49 clone HHSDC22.

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;

KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 OS Homo sapiens.  
 PN WO9839446-A2.

PD 11-SEP-1998.

PD 06-MAR-1998; U04492

PR 07-MAR-1997; US-038621.

PR 07-MAR-1997; US-040161.

PR 07-MAR-1997; US-040162.

PR 07-MAR-1997; US-040163.

PR 07-MAR-1997; US-040333.

PR 07-MAR-1997; US-040334.

PR 07-MAR-1997; US-040336.

PR 11-APR-1997; US-040626.

PR 11-APR-1997; US-043311.

PR 11-APR-1997; US-043312.

PR 11-APR-1997; US-043313.

PR 11-APR-1997; US-043314.

PR 11-APR-1997; US-043315.

PR 11-APR-1997; US-043568.

PR 11-APR-1997; US-043569.

PR 11-APR-1997; US-043576.

PR 11-APR-1997; US-043578.

PR 11-APR-1997; US-043580.

PR 11-APR-1997; US-043669.

PR 11-APR-1997; US-043670.

PR 11-APR-1997; US-043671.

PR 11-APR-1997; US-043672.

PR 11-APR-1997; US-043674.

PR 23-MAY-1997; US-047492.

PR 23-MAY-1997; US-047500.

PR 23-MAY-1997; US-047501.

PR 23-MAY-1997; US-047502.

PR 23-MAY-1997; US-047503.

PR 23-MAY-1997; US-047581.

PR 23-MAY-1997; US-047582.

PR 23-MAY-1997; US-047583.

PR 23-MAY-1997; US-047584.

PR 23-MAY-1997; US-047585.

PR 23-MAY-1997; US-047586.

PR 23-MAY-1997; US-047587.

PR 23-MAY-1997; US-047588.

PR 23-MAY-1997; US-047589.

PR 23-MAY-1997; US-047590.

PR 23-MAY-1997; US-047592.

PR 23-MAY-1997; US-047593.

PR 23-MAY-1997; US-047594.

PR 23-MAY-1997; US-047595.

PR 23-MAY-1997; US-047596.

PR 23-MAY-1997; US-047597.

PR 23-MAY-1997; US-047598.

PR 23-MAY-1997; US-047599.

PR 23-MAY-1997; US-047600.

PR 23-MAY-1997; US-047601.

PR 23-MAY-1997; US-047612.

PR 23-MAY-1997; US-047613.

PR 23-MAY-1997; US-047614.

PR 23-MAY-1997; US-047615.

PR 23-MAY-1997; US-047617.

PR 23-MAY-1997; US-047618.

PR 23-MAY-1997; US-047632.

PR 06-JUN-1997; US-047633.

PR 06-JUN-1997; US-048964.

PR 22-AUG-1997; US-056630.

PR 22-AUG-1997; US-056631.

PR 22-AUG-1997; US-056632.

PR 22-AUG-1997; US-056636.

PR 22-AUG-1997; US-056637.  
PR 22-AUG-1997; US-056662.  
PR 22-AUG-1997; US-056664.  
PR 22-AUG-1997; US-056845.  
PR 22-AUG-1997; US-056862.  
PR 22-AUG-1997; US-056864.  
PR 22-AUG-1997; US-056872.  
PR 22-AUG-1997; US-056874.  
PR 22-AUG-1997; US-056875.  
PR 22-AUG-1997; US-056876.  
PR 22-AUG-1997; US-056877.  
PR 22-AUG-1997; US-056878.  
PR 22-AUG-1997; US-056879.  
PR 22-AUG-1997; US-056880.  
PR 22-AUG-1997; US-056881.  
PR 22-AUG-1997; US-056882.  
PR 22-AUG-1997; US-056884.  
PR 22-AUG-1997; US-056886.  
PR 22-AUG-1997; US-056887.  
PR 22-AUG-1997; US-056888.  
PR 22-AUG-1997; US-056889.  
PR 22-AUG-1997; US-056892.  
PR 22-AUG-1997; US-056893.  
PR 22-AUG-1997; US-056894.  
PR 22-AUG-1997; US-056903.  
PR 22-AUG-1997; US-056908.  
PR 22-AUG-1997; US-056909.  
PR 22-AUG-1997; US-056910.  
PR 22-AUG-1997; US-056911.  
PR 05-SEP-1997; US-057650.  
PR 05-SEP-1997; US-057761.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,  
PI Feng P, Ferlie AM, Fischer CL, Graves KA, Greene JM, Hu JS,  
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,  
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;  
DR WPI: 98-609887/51.  
DR P-FSDB; W75105.  
PT New isolated human genes and the secreted polypeptides they encode  
PT - useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders  
PS Claim 1: Page 205-206; 447pp; English.  
CC This sequence represents a nucleic acid molecule which encodes a secreted  
CC human protein. The gene number, and the clone it is derived from, are  
CC detailed in the descriptor line. The gene can be used to generate fusion  
CC proteins by linking to the gene to a human immunoglobulin Fc portion  
CC (e.g. V34145) for increasing the stability of the fused protein as  
CC compared to the human protein only.  
CC The invention relates to 70 novel genes and their fragments (nucleic acid  
CC sequences: V34154-V34276; amino acid sequences W75057-W75179) which  
CC are useful for preventing, treating or ameliorating medical conditions  
CC e.g. by protein or gene therapy. Also, pathological conditions can be  
CC diagnosed by determining the amount of the new polypeptides in a sample  
CC or by determining the presence of mutations in the new polynucleotides.  
CC Specific uses are described for each of the 70 polynucleotides, based on  
CC which tissues they are most highly expressed in (see V34154 for described  
CC uses).  
SQ Sequence 428 BP; 99 A; 65 C; 107 G; 153 T;

Query Match 1.3%; Score 18; DB 1; Length 428;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1381 atttgaaaaaa 1398  
|||||  
Db 382 ATTTCGAAAAA 399

Search completed: May 1, 2000, 18:50:59  
Job time: 18743 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:01:51 ; Search time 181.17 Seconds  
(without alignments)  
923.845 Million cell updates/sec

Title: US-09-215-435-71  
Perfect score: 1398  
Sequence: 1 gatcgctgagggcccgagc.....acatttgaaaaaaaaaaaaa 1398

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 214294 seqs, 59861574 residues

Word size : 0

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PTUS9\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	1.4	855	7 5185441-40	Patent No. 5185441
2	20	1.4	855	7 5223394-3	Patent No. 5223394
3	20	1.4	863	1 US-07-940-861-11	Sequence 11, Appl
4	20	1.4	863	2 US-08-459-512-11	Sequence 11, Appl
5	20	1.4	863	4 US-08-459-657-11	Sequence 11, Appl
6	20	1.4	863	4 US-08-460-132-11	Sequence 11, Appl
7	20	1.4	863	6 PCT-US92-02050-11	Sequence 11, Appl
8	20	1.4	863	7 5185441-35	Patent No. 5185441
9	20	1.4	863	7 5223394-5	Patent No. 5223394
10	20	1.4	1009	7 5223394-8	Patent No. 5223394
11	20	1.4	1078	7 5223394-10	Patent No. 5223394
12	19	1.4	1268	2 US-08-910-973-12	Sequence 12, Appl
13	19	1.4	2245	1 US-08-203-905B-1	Sequence 1, Appl
14	18	1.3	115	1 US-08-120-827-88	Sequence 88, Appl
15	18	1.3	115	2 US-08-478-675-88	Sequence 88, Appl
16	18	1.3	199	2 US-08-330-108-4	Sequence 4, Appl
17	18	1.3	199	6 PCT-US92-10087-4	Sequence 4, Appl
18	18	1.3	594	4 US-08-289-458-1	Sequence 1, Appl
19	18	1.3	594	4 US-08-761-549-1	Sequence 1, Appl
20	18	1.3	880	2 US-08-616-368A-7	Sequence 7, Appl
21	18	1.3	1023	2 US-08-698-551-7	Sequence 7, Appl
22	18	1.3	1023	3 US-08-602-228-7	Sequence 7, Appl
23	18	1.3	1023	3 US-08-649-341A-7	Sequence 7, Appl
24	18	1.3	1023	3 US-08-494-440B-7	Sequence 7, Appl
25	18	1.3	1023	3 US-08-533-901B-7	Sequence 7, Appl
26	18	1.3	1023	3 US-08-839-032A-7	Sequence 7, Appl

```

27 18 1.3 1023 4 US-08-839-031A-7 Sequence 7, Appl
28 18 1.3 1023 6 PCT-US92-01196-5 Sequence 5, Appl
29 18 1.3 1023 6 PCT-US95-12724-7 Sequence 7, Appl
30 18 1.3 1045 4 US-09-014-969-6 Sequence 6, Appl
31 18 1.3 1233 1 US-08-289-458-4 Sequence 4, Appl
32 18 1.3 1233 4 US-08-761-549-4 Sequence 4, Appl
33 18 1.3 1727 1 US-08-289-458-3 Sequence 3, Appl
34 18 1.3 1727 4 US-08-761-549-3 Sequence 3, Appl
35 18 1.3 1860 4 US-08-372-652-7 Sequence 7, Appl
36 18 1.3 1860 6 PCT-US95-16311-7 Sequence 7, Appl
37 18 1.3 2022 3 US-08-937-540-7 Sequence 7, Appl
38 18 1.3 2377 3 US-08-967-101-26 Sequence 26, Appl
39 18 1.3 2377 4 US-08-592-541-26 Sequence 26, Appl
40 18 1.3 2452 1 US-07-756-250-15 Sequence 15, Appl
41 18 1.3 3138 1 US-07-867-106-4 Sequence 4, Appl
42 18 1.3 3157 7 5198347-3 Patent No. 5198347
43 18 1.3 3171 4 US-08-868-786-5 Sequence 5, Appl
44 18 1.3 5532 4 US-08-475-035-3 Sequence 3, Appl
45 18 1.3 246240 3 US-08-724-394A-20 Sequence 20, Appl

```

## ALIGNMENTS

```

RESULT 1
5185441-40
; Patent No. 5185441
; APPLICANT: WALLNER, BARBARA P.; HESSIONS, CATHERINE
; TITLE OF INVENTION: DNA SEQUENCES, RECOMBINANT DNA
; MOLECULES AND PROCESSES FOR PRODUCING PI-LINKED LYMPHOCYTE
; FUNCTION ASSOCIATED ANTIGEN-3
; NUMBER OF SEQUENCES: 41
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/237,309
; FILING DATE: 26-AUG-1988
; SEQ ID NO: 40:
; LENGTH: 855
5185441-40

```

```

Query Match 1.4%; Score 20; DB 7; Length 855;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1379 acatttgaaaaaaaaaaaaa 1398
Db 814 acatttgaaaaaaaaaaaaa 833

```

```

RESULT 2
5223394-3
; Patent No. 5223394
; APPLICANT: WALLNER, BARBARA
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULE COMPRISING
; LYMPHOCYTE FUNCTION-ASSOCIATED ANTIGEN 3 PHOSPHATIDYLINOSITOL
; LINKAGE SIGNAL SEQUENCE
; NUMBER OF SEQUENCES: 12
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/335,688
; FILING DATE: 10-APR-1989
; SEQ ID NO: 3:
; LENGTH: 855
5223394-3

```

```

Query Match 1.4%; Score 20; DB 7; Length 855;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1379 acatttgaaaaaaaaaaaaa 1398
Db 814 acatttgaaaaaaaaaaaaa 833

```



```
; TITLE OF INVENTION: FUNCTION ASSOCIATED ANTIGEN 3
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 875 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10022-6250
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/459,657
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/02050
; FILING DATE: 12-MAR-1992
; APPLICATION NUMBER: US 07/667,971
; FILING DATE: 12-MAR-1991
; APPLICATION NUMBER: US 07/770,967
; FILING DATE: 07-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: HALEY, James F., Jr.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: B151CIP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)715-0600
; TELEFAX: (212)715-0673
; TELEX: 14-8367
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 863 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 18..737
; US-08-459-657-11

Query Match 1.4%; Score 20; DB 4; Length 863;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1379 acatttgaaaaa 1398
DB 822 ACATTGAAAAA 841

RESULT 6
PCT-US92-02050-11
; Sequence 11, Application US/08460132
; Patent No. 5928643
; GENERAL INFORMATION:
; APPLICANT: BIOGEN, INC.
; APPLICANT: WALLNER, Barbara P.
; APPLICANT: MILLER, Glenn T.
; APPLICANT: ROSA, Margaret D.
; TITLE OF INVENTION: CD2-BINDING DOMAIN OF LYMPHOCYTE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 875 Third Avenue
; CITY: New York
```

```
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10022-6250
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/460,132
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/940,861
; FILING DATE: 21-OCT-1992
; APPLICATION NUMBER: PCT/US92/02050
; FILING DATE: 12-MAR-1992
; APPLICATION NUMBER: US 07/667,971
; FILING DATE: 12-MAR-1991
; APPLICATION NUMBER: US 07/770,967
; FILING DATE: 07-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: HALEY, James F., Jr.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: B151CIP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)715-0600
; TELEFAX: (212)715-0673
; TELEX: 14-8367
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 863 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 18..737
; US-08-460-132-11

Query Match 1.4%; Score 20; DB 4; Length 863;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1379 acatttgaaaaa 1398
DB 822 ACATTGAAAAA 841

RESULT 7
PCT-US92-02050-11
; Sequence 11, Application PC/TUS9202050
; GENERAL INFORMATION:
; APPLICANT: BIOGEN, INC.
; APPLICANT: WALLNER, Barbara P.
; APPLICANT: MILLER, Glenn T.
; APPLICANT: ROSA, Margaret D.
; TITLE OF INVENTION: CD2-BINDING DOMAIN OF LYMPHOCYTE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 875 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10022-6250
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/02050  
FILING DATE: 19920312  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/667,971  
FILING DATE: 12-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/770,967  
FILING DATE: 07-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: HALEY, James F., Jr.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B151CIP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)715-0600  
TELEFAX: (212)715-0673  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 863 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
HYPOHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 18..737  
PCT-US92-02050-11

Query Match 1.4%; Score 20; DB 6; Length 863;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1379 acatttgaaaaaa 1398  
|||||  
Db 822 ACATTGAAAAA 841

RESULT 8  
5185441-35  
; Patent No. 5185441  
; APPLICANT: WALLNER, BARBARA P.; HESSIONS, CATHERINE  
; TITLE OF INVENTION: DNA SEQUENCES, RECOMBINANT DNA  
; MOLECULES AND PROCESSES FOR PRODUCING PI-LINKED LYMPHOCYTE  
; FUNCTION ASSOCIATED ANTIGEN-3  
; NUMBER OF SEQUENCES: 41  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/237,309  
; FILING DATE: 26-AUG-1988  
; SEQ ID NO:35:  
; LENGTH: 863  
5185441-35

Query Match 1.4%; Score 20; DB 7; Length 863;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1379 acatttgaaaaaa 1398  
|||||  
Db 822 acatttgaaaaaa 841

RESULT 9  
5223394-5  
; Patent No. 5223394

; APPLICANT: WALLNER, BARBARA  
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULE COMPRISING  
; LYMPHOCYTE FUNCTION-ASSOCIATED ANTIGEN 3 PHOSPHATIDYLINOSITOL  
; LINKAGE SIGNAL SEQUENCE  
; NUMBER OF SEQUENCES: 12  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/335,688  
; FILING DATE: 10-APR-1989  
; SEQ ID NO:5:  
; LENGTH: 863  
5223394-5

Query Match 1.4%; Score 20; DB 7; Length 863;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1379 acatttgaaaaaa 1398  
|||||  
Db 822 acatttgaaaaaa 841

RESULT 10  
5223394-8  
; Patent No. 5223394  
; APPLICANT: WALLNER, BARBARA  
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULE COMPRISING  
; LYMPHOCYTE FUNCTION-ASSOCIATED ANTIGEN 3 PHOSPHATIDYLINOSITOL  
; LINKAGE SIGNAL SEQUENCE  
; NUMBER OF SEQUENCES: 12  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/335,688  
; FILING DATE: 10-APR-1989  
; SEQ ID NO:8:  
; LENGTH: 1009  
5223394-8

Query Match 1.4%; Score 20; DB 7; Length 1009;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1379 acatttgaaaaaa 1398  
|||||  
Db 970 acatttgaaaaaa 989

RESULT 11  
5223394-10  
; Patent No. 5223394  
; APPLICANT: WALLNER, BARBARA  
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULE COMPRISING  
; LYMPHOCYTE FUNCTION-ASSOCIATED ANTIGEN 3 PHOSPHATIDYLINOSITOL  
; LINKAGE SIGNAL SEQUENCE  
; NUMBER OF SEQUENCES: 12  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/335,688  
; FILING DATE: 10-APR-1989  
; SEQ ID NO:10:  
; LENGTH: 1078  
5223394-10

Query Match 1.4%; Score 20; DB 7; Length 1078;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1379 acatttgaaaaaa 1398  
|||||  
Db 1039 acatttgaaaaaa 1058

RESULT 12

```
US-08-910-973-12/c
; Sequence 12, Application US/08910973
; Patent No. 5795723
; GENERAL INFORMATION:
; APPLICANT: Tapscott, Stephen J.
; APPLICANT: Olson, James M.
; TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,973
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION: 435
; APPLICATION NUMBER: US 08/239,238
; FILING DATE: 06-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17532
; FILING DATE: 30-October-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: EHC-1-10958
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100; 206-224-0735 (direct)
; TELEFAX: 206-225-0779
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1268 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: 20A1 (neurod3)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..768
US-08-910-973-12

Query Match 1.4%; Score 19; DB 2; Length 1268;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 734 gggctcagagcggggagg 752
|||||
DB 650 GGGTCAGAGCGGGGAGG 632

RESULT 13
US-08-203-905B-1
; Sequence 1, Application US/08203905B
; Patent No. 5646249
; GENERAL INFORMATION:
; APPLICANT: KAYE, FEDERIC J.
; APPLICANT: OTTERSON, GREGORY A.
```

```
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A
; TITLE OF INVENTION: NOVEL CHAPERONE PROTEIN
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR
; STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
; CITY: NEWPORT BEACH
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/203,905B
; FILING DATE: February 28, 1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: KIRKPATRICK, ANITA M.
; REGISTRATION NUMBER: 32,617
; REFERENCE/DOCKET NUMBER: NIH089.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-235-8550
; TELEFAX: 619-235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2245 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 37..1452
US-08-203-905B-1

Query Match 1.4%; Score 19; DB 1; Length 2245;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1380 cattgaaaaaiaaaaaa 1398
|||||
DB 2221 CATTGAAAAAIAAAAAA 2239

RESULT 14
US-08-120-827-88
; Sequence 88, Application US/08120827
; Patent No. 5525495
; GENERAL INFORMATION:
; APPLICANT: KEENE, JACK D.
; APPLICANT: KING, PETER H.
; APPLICANT: LEVINE, TODD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

us-09-215-435-71.olig.rni

Fri May 12 12:27:58 2000

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/120,827
; FILING DATE: 15-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NO. 5525495man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 714-158-0 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)413-3000
; TELEFAX: (703)413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 115 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
; US-08-120-827-88

Query Match      1.3%; Score 18; DB 2; Length 115;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1381 attgaaaaaa 1398
Db 97 ATTGAAAAAA 114

Search completed: May 1, 2000, 16:01:59
Job time: 18184 sec

```

```

Query Match      1.3%; Score 18; DB 1; Length 115;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1381 attgaaaaaa 1398
Db 97 ATTGAAAAAA 114

```

```

RESULT 15
US-08-478-675-88
; Sequence 88, Application US/08478675
; Patent No. 5773246
; GENERAL INFORMATION:
; APPLICANT: KEENE, JACK D.
; APPLICANT: KING, PETER H.
; APPLICANT: LEVINE, TODD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,675
; FILING DATE: 07-JUN-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/120,827
; FILING DATE: 15-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5773246man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 714-158-0 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)413-3000
; TELEFAX: (703)413-2220
; TELEX: 248855 OPAT UR

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 19:59:20 ; Search time 4088.29 Seconds  
(without alignments)  
-1172.629 Million cell updates/sec

Title: US-09-215-435-71

Perfect score: 1398

Sequence: 1 gatccttgaggcccccagc.....acatttgaaaaaaaaaaaaa 1398

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 5765378 seqs, -1714608768 residues

Word size : 0

Total number of hits satisfying chosen parameters: 10530736

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

Pending\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/1/pna/PCTUS9\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pna/US081A\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pna/US081B\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pna/US081C\_COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pna/US081D\_COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/pna/US082A\_COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pna/US082B\_COMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pna/US082C\_COMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pna/US083A\_COMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pna/US083B\_COMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pna/US084A\_COMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pna/US084B\_COMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pna/US084C\_COMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq:\*
- 17: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq:\*
- 18: /cgn2\_6/ptodata/1/pna/US087A\_COMB.seq:\*
- 19: /cgn2\_6/ptodata/1/pna/US087B\_COMB.seq:\*
- 20: /cgn2\_6/ptodata/1/pna/US087C\_COMB.seq:\*
- 21: /cgn2\_6/ptodata/1/pna/US088A\_COMB.seq:\*
- 22: /cgn2\_6/ptodata/1/pna/US088B\_COMB.seq:\*
- 23: /cgn2\_6/ptodata/1/pna/US088C\_COMB.seq:\*
- 24: /cgn2\_6/ptodata/1/pna/US088D\_COMB.seq:\*
- 25: /cgn2\_6/ptodata/1/pna/US089A\_COMB.seq:\*
- 26: /cgn2\_6/ptodata/1/pna/US089B\_COMB.seq:\*
- 27: /cgn2\_6/ptodata/1/pna/US089C\_COMB.seq:\*
- 28: /cgn2\_6/ptodata/1/pna/US089D\_COMB.seq:\*
- 29: /cgn2\_6/ptodata/1/pna/US089E\_COMB.seq:\*
- 30: /cgn2\_6/ptodata/1/pna/US089F\_COMB.seq:\*
- 31: /cgn2\_6/ptodata/1/pna/US089G\_COMB.seq:\*
- 32: /cgn2\_6/ptodata/1/pna/US089H\_COMB.seq:\*
- 33: /cgn2\_6/ptodata/1/pna/US089I\_COMB.seq:\*
- 34: /cgn2\_6/ptodata/1/pna/US090A\_COMB.seq:\*
- 35: /cgn2\_6/ptodata/1/pna/US090B\_COMB.seq:\*
- 36: /cgn2\_6/ptodata/1/pna/US090C\_COMB.seq:\*
- 37: /cgn2\_6/ptodata/1/pna/US090D\_COMB.seq:\*
- 38: /cgn2\_6/ptodata/1/pna/US090E\_COMB.seq:\*
- 39: /cgn2\_6/ptodata/1/pna/US091A\_COMB.seq:\*
- 40: /cgn2\_6/ptodata/1/pna/US091B\_COMB.seq:\*
- 41: /cgn2\_6/ptodata/1/pna/US091C\_COMB.seq:\*
- 42: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq:\*
- 43: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq:\*

44: /cgn2\_6/ptodata/1/pna/US092C\_COMB.seq:\*

45: /cgn2\_6/ptodata/1/pna/US092D\_COMB.seq:\*

46: /cgn2\_6/ptodata/1/pna/US092E\_COMB.seq:\*

47: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq:\*

48: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq:\*

49: /cgn2\_6/ptodata/1/pna/US093C\_COMB.seq:\*

50: /cgn2\_6/ptodata/1/pna/US093D\_COMB.seq:\*

51: /cgn2\_6/ptodata/1/pna/US093E\_COMB.seq:\*

52: /cgn2\_6/ptodata/1/pna/US093F\_COMB.seq:\*

53: /cgn2\_6/ptodata/1/pna/US094A\_COMB.seq:\*

54: /cgn2\_6/ptodata/1/pna/US094B\_COMB.seq:\*

55: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq:\*

56: /cgn2\_6/ptodata/1/pna/US6001A\_COMB.seq:\*

57: /cgn2\_6/ptodata/1/pna/US6001B\_COMB.seq:\*

58: /cgn2\_6/ptodata/1/pna/US6002A\_COMB.seq:\*

59: /cgn2\_6/ptodata/1/pna/US6002B\_COMB.seq:\*

60: /cgn2\_6/ptodata/1/pna/US6002C\_COMB.seq:\*

61: /cgn2\_6/ptodata/1/pna/US6003A\_COMB.seq:\*

62: /cgn2\_6/ptodata/1/pna/US6003B\_COMB.seq:\*

63: /cgn2\_6/ptodata/1/pna/US6004A\_COMB.seq:\*

64: /cgn2\_6/ptodata/1/pna/US6004B\_COMB.seq:\*

65: /cgn2\_6/ptodata/1/pna/US6004C\_COMB.seq:\*

66: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq:\*

67: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq:\*

68: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq:\*

69: /cgn2\_6/ptodata/1/pna/US6008A\_COMB.seq:\*

70: /cgn2\_6/ptodata/1/pna/US6008B\_COMB.seq:\*

71: /cgn2\_6/ptodata/1/pna/US6008C\_COMB.seq:\*

72: /cgn2\_6/ptodata/1/pna/US6009A\_COMB.seq:\*

73: /cgn2\_6/ptodata/1/pna/US6009B\_COMB.seq:\*

74: /cgn2\_6/ptodata/1/pna/US6010A\_COMB.seq:\*

75: /cgn2\_6/ptodata/1/pna/US6010B\_COMB.seq:\*

76: /cgn2\_6/ptodata/1/pna/US6011A\_COMB.seq:\*

77: /cgn2\_6/ptodata/1/pna/US6011B\_COMB.seq:\*

78: /cgn2\_6/ptodata/1/pna/US6012A\_COMB.seq:\*

79: /cgn2\_6/ptodata/1/pna/US6012B\_COMB.seq:\*

80: /cgn2\_6/ptodata/1/pna/US6013A\_COMB.seq:\*

81: /cgn2\_6/ptodata/1/pna/US6013B\_COMB.seq:\*

82: /cgn2\_6/ptodata/1/pna/US6014A\_COMB.seq:\*

83: /cgn2\_6/ptodata/1/pna/US6014B\_COMB.seq:\*

84: /cgn2\_6/ptodata/1/pna/US6014C\_COMB.seq:\*

85: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq:\*

86: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq:\*

87: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq:\*

88: /cgn2\_6/ptodata/2/pna/PCT\_NEW\_COMB.seq:\*

89: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB.seq:\*

90: /cgn2\_6/ptodata/2/pna/US07\_NEW\_COMB.seq:\*

91: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq:\*

92: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:\*

93: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1398	100.0	1398	42	US-09-215-435-71
2	900	64.4	1812	54	US-09-471-275-9883
3	751	53.7	1423	73	US-60-096-116-60
4	629	45.0	1285	42	US-09-215-435-70
5	618	44.2	1174	42	US-09-215-435-69
6	618	44.2	1197	73	US-60-096-116-58
7	578	41.3	1309	73	US-60-096-116-59
8	536	38.3	899	87	US-60-172-377-10510
9	536	38.3	899	93	US-60-172-377-10510
10	536	38.3	899	93	US-60-172-377-10510
11	448	32.0	1451	54	US-09-471-275-1784
12	436	31.2	502	44	US-09-277-227-17958
13	436	31.2	502	49	US-09-346-956-16218

```

14 435 31.1 1331 51 US-09-371-168-6486
15 431 30.8 495 48 US-09-332-782-2467
16 418 29.9 458 44 US-09-277-227-17550
17 418 29.9 458 49 US-09-346-956-15990
18 399 27.8 1548 87 US-60-172-360-14738
19 384 27.5 519 93 US-60-185-215-3314
20 377 27.0 432 92 US-09-489-036-32457
21 354 25.3 456 48 US-09-332-782-977
22 304 21.7 458 50 US-09-353-690-6547
23 288 20.6 385 44 US-09-274-861-5090
24 288 20.6 413 44 US-09-274-861-5089
25 282 20.2 282 25 US-08-901-904-65
26 279 20.0 444 49 US-09-346-956-9764
27 275 19.7 401 44 US-09-271-490-387
28 255 18.2 509 46 US-09-293-972-6612
29 249 17.8 364 48 US-09-332-782-607
30 249 17.8 364 92 US-09-515-694-607
31 236 16.9 407 92 US-09-489-036-27213
32 231 16.5 420 92 US-09-489-036-385
33 230 16.5 400 44 US-09-277-227-8298
34 230 16.5 400 92 US-09-240-371-3363
35 228 16.3 275 25 US-08-901-904-1384
36 224 16.0 1077 42 US-09-205-070-12701
37 224 16.0 1077 49 US-09-340-623-12701
38 222 15.9 375 92 US-09-489-036-15427
39 220 15.7 396 46 US-09-293-972-11984
40 214 15.3 214 28 US-08-940-864-456
41 214 15.3 214 60 US-06-027-236-456
42 195 13.9 444 45 US-09-287-618-9354
43 190 13.6 417 92 US-09-489-036-34133
44 187 13.4 468 45 US-09-289-768-30370
45 187 13.4 468 48 US-09-332-782-27576

```

# ALIGNMENTS

```

RESULT 1
US-09-215-435-71
; GENERAL INFORMATION: US/09215435A
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric
; TITLE OF INVENTION: Extended cDNAs
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215,435A
; EARLIER FILING DATE: 1998-12-17
; EARLIER FILING DATE: 1997-12-17
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 71
; LENGTH: 1398
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1171
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 2..205
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 7.3
; FEATURE:
; NAME/KEY: polyA_signal

```

```

; LOCATION: 1368..1373
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 1386..1398
US-09-215-435-71

Query Match      100.0%; Score 1398; DB 42; Length 1398;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gatgcctgagggcccgagcagctcacctggccagcagatttgtgaatgagggcctgagggc 60
Db 1 gatgcctgagggcccgagcagctcacctggccagcagatttgtgaatgagggcctgagggc 60
QY 61 gctgggttcggcggtcgctggagaagtcctctgtcagcccaacccctgaggtgcctt 120
Db 61 gctgggttcggcggtcgctggagaagtcctctgtcagcccaacccctgaggtgcctt 120
QY 121 tgagagcagtgctaccgcagctcatctcagcttcagcccgcggaaggagctgcctgatact 180
Db 121 tgagagcagtgctaccgcagctcatctcagcttcagcccgcggaaggagctgcctgatact 180
QY 181 gagccctctgcctggggccagcccaacacagagcaactggccctggtcttcgcttcgg 240
Db 181 gagccctctgcctggggccagcccaacacagagcaactggccctggtcttcgcttcgg 240
QY 241 catgtccgctcttttcagctggtgcccgcgagagagctgccacgcatgccacacctgcg 300
Db 241 catgtccgctcttttcagctggtgcccgcgagagagctgccacgcatgccacacctgcg 300
QY 301 cttttacagggcccgccctggccccggctgcctctatatttcgtggacaatccgcggtt 360
Db 301 cttttacagggcccgccctggccccggctgcctctatatttcgtggacaatccgcggtt 360
QY 361 cggccgctggacctgggggaaagtggcagccggcgccggccctgtgtcttcgagga 420
Db 361 cggccgctggacctgggggaaagtggcagccggcgccggccctgtgtcttcgagga 420
QY 421 gtaccagcagttcagggagaatgtctacgaaacctagcggataagcgttttgaccggcc 480
Db 421 gtaccagcagttcagggagaatgtctacgaaacctagcggataagcgttttgaccggcc 480
QY 481 catctgcgagggccctcctggaccagaggttcttcaatggcattggccaactatctgcgggc 540
Db 481 catctgcgagggccctcctggaccagaggttcttcaatggcattggccaactatctgcgggc 540
QY 541 agagatcctgtaccggctgaagatcccccccttgagaagggccgctcggtctctggaggc 600
Db 541 agagatcctgtaccggctgaagatcccccccttgagaagggccgctcggtctctggaggc 600
QY 601 cctgcagcagcacaggccgagcccgagctgacctgagccaggaagataagaccgaagt 660
Db 601 cctgcagcagcacaggccgagcccgagctgacctgagccaggaagataagaccgaagt 660
QY 661 gcagaatccagacctgtggagctatgtcaactcagtgcccaaggaagtgtccagttggg 720
Db 661 gcagaatccagacctgtggagctatgtcaactcagtgcccaaggaagtgtccagttggg 720
QY 721 gggcagaggtctacgggtcagagagcggggagaggacttctgtcctttcgagcctggct 780
Db 721 gggcagaggtctacgggtcagagagcggggagaggacttctgtcctttcgagcctggct 780
QY 781 ggcgtgctatggcatgccagggcatgagctccctgcaggaccggcatggcgctaccatctg 840
Db 781 ggcgtgctatggcatgccagggcatgagctccctgcaggaccggcatggcgctaccatctg 840
QY 841 gttccagggggatcctggaccgttgccacccaaaggccgaagtccgcgaagaaatc 900
Db 841 gttccagggggatcctggaccgttgccacccaaaggccgaagtccgcgaagaaatc 900
QY 901 caaggccacacagctgagtcctcagagacagagtggaggagcgttttgcctcccgagcaggc 960
Db 901 caaggccacacagctgagtcctcagagacagagtggaggagcgttttgcctcccgagcaggc 960

```

```

Db 901 caagggccacacagctgagctcctgagacagagctgagagacgcttgcctccgagcaagcc 960
Qy 961 cccctccaggacacagggcaaaagagagaccttccctgaagagagctgcaacccagcgcc 1020
Db 961 cccctccaggacacagggcaaaagagagaccttccctgaagagagctgcaacccagcgcc 1020
Qy 1021 tgaggggaccagctccagcagagaccagcaagagctccacagctgcccagaagggagag 1080
Db 1021 tgaggggaccagctccagcagagaccagcaagagctccacagctgcccagaagggagag 1080
Qy 1081 gaaggggacagcagcagctcctggccactgcaagaccccggaaggtcaagggctgacatccc 1140
Db 1081 gaaggggacagcagcagctcctggccactgcaagaccccggaaggtcaagggctgacatccc 1140
Qy 1141 atccttgaacacagaggggacacacagctccttagcagagggctccttgcctgcaactca 1200
Db 1141 atccttgaacacagaggggacacacagctccttagcagagggctccttgcctgcaactca 1200
Qy 1201 ccccttctattgcttgcctgcatctgggggtctggaatttttggagcagcgaatatc 1260
Db 1201 ccccttctattgcttgcctgcatctgggggtctggaatttttggagcagcgaatatc 1260
Qy 1261 tgaaggtcaaacagcagcagctcctgctgcttccctgcacaaactcctatgttttaattgtac 1320
Db 1261 tgaaggtcaaacagcagcagctcctgctgcttccctgcacaaactcctatgttttaattgtac 1320
Qy 1321 cccattctccacatctttaaagctcatgtgaaataatgctgcaatttttaataaactgatac 1380
Db 1321 cccattctccacatctttaaagctcatgtgaaataatgctgcaatttttaataaactgatac 1380
Qy 1381 atttgaaaaaataaaaaa 1398
Db 1381 atttgaaaaaataaaaaa 1398

```

RESULT 2

```

US-09-471-275-9883
; Sequence 9883, Application US/09471275
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Contigs Obtained
; FROM VARIOUS LIBRARIES
; FILE REFERENCE: 782
; CURRENT APPLICATION NUMBER: US/09/471,275
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: US 09/235,076
; EARLIER FILING DATE: 1999-01-20
; EARLIER APPLICATION NUMBER: US 09/234,611
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: US 09/240,371
; EARLIER FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: US 09/277,227
; EARLIER FILING DATE: 1999-03-25
; EARLIER APPLICATION NUMBER: US 09/271,490
; EARLIER FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: US 09/293,972
; EARLIER FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: US 09/274,861
; EARLIER FILING DATE: 1999-03-23
; EARLIER APPLICATION NUMBER: US 60/125,453
; EARLIER FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: US 60/126,605
; EARLIER FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: US 09/306,350
; EARLIER FILING DATE: 1999-05-07
; EARLIER APPLICATION NUMBER: US 09/399,720
; EARLIER FILING DATE: 1999-09-21
; EARLIER APPLICATION NUMBER: US 09/404,284
; EARLIER FILING DATE: 1999-09-21
; EARLIER APPLICATION NUMBER:
; EARLIER FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 10451
; SOFTWARE: pt_CT_genes Version 1.0

```

```

; SEQ ID NO 9883
; LENGTH: 1812
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1721)...(0)
; OTHER INFORMATION: similar to g1214042 in the genepept database release 114,
; OTHER INFORMATION: Run with FASTX 3.3t00, default parameters
US-09-471-275-9883

```

```

Query Match 64.48; Score 900; DB 54; Length 1812;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 950; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gatgctgagggcccccagctgcaacctggccagccagttgtgaatgagggcctgcagggc 60
Db 853 gatgctgagggcccccagctgcaacctggccagccagttgtgaatgagggcctgcagggc 912
Qy 61 gctggtgttcggcgctgctgaggaagctcctctgtagccgcaacccctgaggtgccctt 120
Db 913 gctggtgttcggcgctgctgaggaagctcctctgtagccgcaacccctgaggtgccctt 972
Qy 121 tgagagcagtgctacacgcatctcagcttccgcccggcgaagagctgcgctgatact 180
Db 973 tgagagcagtgctacacgcatctcagcttccgcccggcgaagagctgcgctgatact 1032
Qy 181 gagccctctgctggggcccaacagcagcagcagcagcagcagcagcagcagcagcagc 240
Db 1033 gagccctctgctggggcccaacagcagcagcagcagcagcagcagcagcagcagcagc 1092
Qy 241 catgtccggctcttttcagctggtgcccgcgagagctgcacgcccacacacacacacac 300
Db 1093 catgtccggctcttttcagctggtgcccgcgagagctgcacgcccacacacacacac 1152
Qy 301 attttacacggcccgccctgcccgcgctgcctcctatgtttctgtagacatccgcggtt 360
Db 1153 attttacacggcccgccctgcccgcgctgcctcctatgtttctgtagacatccgcggtt 1212
Qy 361 cggccgctggacacctggggaaagtgacacggcgccgcccggcggcggcggcggcggcggcgg 420
Db 1213 cggccgctggacacctggggaaagtgacacggcgccgcccggcggcggcggcggcggcgg 1272
Qy 421 gtaccagcagttcagggagaaatgtctacgaaacacacacacacacacacacacacacac 480
Db 1273 gtaccagcagttcagggagaaatgtctacgaaacacacacacacacacacacacacac 1332
Qy 481 catctgcagggccctcctcctggaccagaggtttcttaattggcattggcaactatctgcggc 540
Db 1333 catctgcagggccctcctcctggaccagaggtttcttaattggcattggcattatctgcggc 1392
Qy 541 agagatcctgacggctgaaagatcccccctttgagagggcccgctggcttgcagggc 600
Db 1393 agagatcctgacggctgaaagatcccccctttgagagggcccgctggcttgcagggc 1452
Qy 601 cctgcagcagcacagcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc 660
Db 1453 cctgcagcagcacagcggcggcggcggcggcggcggcggcggcggcggcggcggcggcgg 1512
Qy 661 gaagaatccagacctgctggagctatgtcactcagtgcccaggaggtgggtccagttggg 720
Db 1513 gaagaatccagacctgctggagctatgtcactcagtgcccaggaggtgggtccagttggg 1572
Qy 721 gggcagaggtacgggtcagagagcggggagagaggaacttgcgcttgcagcgtgct 780
Db 1573 gggcagaggtacgggtcagagagcggggagagaggaacttgcgcttgcagcgtgct 1632
Qy 781 ggcgtgctatggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 840
Db 1633 ggcgtgctatggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1692
Qy 841 gttccaggggatacctgacgcttggccaccacaaaggcgcaagtcccccaaaagaatac 900

```







```
Db 181 gagccctctgctggggccacagcctcaacagagagagcactggccctgggtcttccgcttggg 240
QY 241 catgtccggctcttttcagctggtgccccgcgagagagctgcccacccatgccacctggg 300
Db 241 catgtccggctcttttcagctggtgccccgcgagagagctgcccacccatgccacctggg 300
QY 301 cttttacacggcccgctgccccgcgctcgccctatgttttcgtagacatccgcggtt 360
Db 301 cttttacacggcccgctgccccgcgctcgccctatgttttcgtagacatccgcggtt 360
QY 361 cggcgctgggacctgggggaaagtggcagccggcgccggccctgtgtcttcaggga 420
Db 361 cggcgctgggacctgggggaaagtggcagccggcgccggccctgtgtcttcaggga 420
QY 421 gtaccagcagttcaggaggaatgtctacgaaacctagcggataaggccctttgacggcc 480
Db 421 gtaccagcagttcaggaggaatgtctacgaaacctagcggataaggccctttgacggcc 480
QY 481 catctgcagggccctcctggacacagaggttctcaatggcattgcaactatctcgggc 540
Db 481 catctgcagggccctcctggacacagaggttctcaatggcattgcaactatctcgggc 540
QY 541 agagatcctgtaccggctgaagatccccccctttgagaagcccgctcggtcctggaggc 600
Db 541 agagatcctgtaccggctgaagatccccccctttgagaagcccgctcggtcctggaggc 600
QY 601 cctgcagcagcagccgagcccgagctgacctgacctgagccagagaagataagacaaagt 660
Db 601 cctgcagcagcagccgagcccgagctgacctgacctgagccagagaagataagacaaagt 660
QY 661 gcagaatccagacctgctgagctatgtcaactcagtgccccaaaggagtggtccagtggg 720
Db 661 gcagaatccagacctgctgagctatgtcaactcagtgccccaaaggagtggtccagtggg 720
```

RESULT 6

```
US-60-096-116-58
; Sequence 58, Application US/60096116
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste Dumas Milne
; APPLICANT: Duclert, Aymeric
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SECRETED PROTEINS
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/096.116
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.038PR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1197 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
```

```
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Lymphocytes ganglia
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 2..205
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 7.3
; OTHER INFORMATION: seq LRLILSLPGAQP/QQ
; FEATURE:
; NAME/KEY: poly_a
; LOCATION: 1160..1173
; IDENTIFICATION METHOD: blastn2
US-60-096-116-58

Query Match 44.2%; Score 618; DB 73; Length 1197;
Best Local Similarity 99.7%; Pred. No. 1e-280;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gatgctgagggccccgagctgcacctggccagccagtttgtgaatgaggcctgcaggcc 60
Db 1 GATGCTGAGGGCCCCGAGCTGCACCTGCCAGCCAGTTTGTGAATGAGGCTGCAGGCG 60
QY 61 gctggtgttcggcgctgctggagaaagtctctctcagccgcaacctgagggtgccctt 120
Db 61 GCTGCTGTTCGGCGGCTCGTGGAGAGTCTCTGTCTAGCCGCAACCTGAGGTGCCCTT 120
QY 121 tgagagcagtcctaccgcacatctcagcttcagccgcgcaaggagctgccttgatact 180
Db 121 TGAGAGCAGTCCCTACCCGCACTCTCAGCTTCAGCCGCCGCGCAAGGAGCTGCGCTGATACT 180
QY 181 gagccctctgctggggccccagcggcccaacagggagcactggccctgggtctctccgctcgg 240
Db 181 GAGCCCTCTGCTGGGGCCCCAGCTCAACAGAGGACCTGCGCCCTGTCTTCCGCTTCGG 240
QY 241 catgtccggctcttttcagctggtgccccgcgagagagctgcccacccatgccacctggc 300
Db 241 CATGTCCGGCTCTTTTTCAGCTGGTGCCTCCCGGAGAGAGTGCACCCATGCCCATGCCACTCGG 300
QY 301 cttttacacggcccgctgccccgcgctcgccctatgttttcgtagacatccgcggtt 360
Db 301 CTTTTCACGCGCCCGCTGCCCCCGGCTCGCCCTATGTTTCGTGGACATCCGCGCGTT 360
QY 361 cggcgctgggacctgggggaaagtggcagccggcgccggccctgtgtcttcaggga 420
Db 361 CGGCGCTGGGACCTTGGGGAAAGTGGCAGCGCGGCGCGGCCCTGTGTCTTCAGGA 420
QY 421 gtaccagcagttcaggaggaatgtctacgaaacctagcggataaggccctttgacggcc 480
Db 421 GTACCAGCAGTTCAGGGAGAAATGTCTACGAAACCTAGCGGATAGGCTTTGACCGGCC 480
QY 481 catctgcagggccctcctggaccagaggttctcaatggcattggaactatctcgcggc 540
Db 481 CATCTGCAGGCCCCCTCTGGACCAAGAGGTTCTTCAATGGCATTTGCAACTATCTCGGGGC 540
QY 541 agagatcctgacctggcgtgaagatccccccctttgagaagcccgctcggtcctggaggc 600
Db 541 AGAGATCCTGTACGGGCTGAAGATCCCCCTTTTGAGAAGGCCCGCTCGGCTCTGAGGC 600
QY 601 cctgcagcagcagccgagcccgagctgacctgacctgagccagagaagataagacaaagt 660
Db 601 CCTGCAGCAGCACAGCGCGGAGCCCGGAGCTGACCCCTGAGCCAGAGAGATAAGGACCAAGCT 660
QY 661 gcagaatccagacctgctgagctatgtcaactcagtgccccaaaggagtggtccagtggg 720
Db 661 GCAGAAATTCAGACCTGCTGGAGCTATGTCACTAGTGCCTCAAGGAAGTGGTCCAGTTGGG 720

RESULT 7
US-60-096-116-59
; Sequence 59, Application US/60096116
; GENERAL INFORMATION:
```

APPLICANT: Edwards, Jean-Baptiste Dumas Milne  
APPLICANT: Duclert, Aymeric  
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SECRETED PROTEINS  
NUMBER OF SEQUENCES: 95  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 501 West Broadway  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92101-3505  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Win95  
SOFTWARE: Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/60/096,116  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Israel, Ned A.  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: GENSET.038PR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1309 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: DOUBLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: CDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo Sapiens  
TISSUE TYPE: Lymphocytes ganglia  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 2..205  
IDENTIFICATION METHOD: Von Heijne matrix  
OTHER INFORMATION: score 7.3  
OTHER INFORMATION: seq LRLILSLPLGAQP/QQ  
FEATURE:  
NAME/KEY: poly\_a\_signal  
LOCATION: 1248..1253  
FEATURE:  
NAME/KEY: poly\_a  
LOCATION: 1272..1285  
IDENTIFICATION METHOD: blastn2  
FEATURE:  
NAME/KEY: est  
LOCATION: complement(1149..1267)  
IDENTIFICATION METHOD: blastn2  
OTHER INFORMATION: identity 100  
OTHER INFORMATION: region 10..128  
OTHER INFORMATION: id AA5951925  
FEATURE:  
NAME/KEY: est  
LOCATION: complement(847..1265)  
IDENTIFICATION METHOD: blastn2  
OTHER INFORMATION: identity 99  
OTHER INFORMATION: region 1..419  
OTHER INFORMATION: id AA595778  
FEATURE:  
NAME/KEY: est  
LOCATION: complement(909..1267)  
IDENTIFICATION METHOD: blastn2  
OTHER INFORMATION: identity 99  
OTHER INFORMATION: region 39..396  
OTHER INFORMATION: id AA824509  
FEATURE:  
NAME/KEY: est

LOCATION: complement(912..1267)  
IDENTIFICATION METHOD: blastn2  
OTHER INFORMATION: identity 99  
OTHER INFORMATION: region 33..388  
OTHER INFORMATION: id AA832090  
FEATURE:  
NAME/KEY: est  
LOCATION: complement(985..1267)  
IDENTIFICATION METHOD: blastn2  
OTHER INFORMATION: identity 99  
OTHER INFORMATION: region 24..306  
OTHER INFORMATION: id AA668993  
FEATURE:  
NAME/KEY: est  
LOCATION: complement(756..1267)  
IDENTIFICATION METHOD: blastn2  
OTHER INFORMATION: identity 98  
OTHER INFORMATION: region 10..521  
OTHER INFORMATION: id AA744383  
FEATURE:  
NAME/KEY: est  
LOCATION: complement(767..1267)  
IDENTIFICATION METHOD: blastn2  
OTHER INFORMATION: identity 98  
OTHER INFORMATION: region 12..512  
OTHER INFORMATION: id AA833865  
FEATURE:  
NAME/KEY: est  
LOCATION: complement(771..1267)  
IDENTIFICATION METHOD: blastn2  
OTHER INFORMATION: identity 98  
OTHER INFORMATION: region 10..504  
OTHER INFORMATION: id AA831660  
FEATURE:  
NAME/KEY: est  
LOCATION: complement(788..1265)  
IDENTIFICATION METHOD: blastn2  
OTHER INFORMATION: identity 98  
OTHER INFORMATION: region 1..478  
OTHER INFORMATION: id AA604501  
FEATURE:  
NAME/KEY: est  
LOCATION: complement(845..1265)  
IDENTIFICATION METHOD: blastn2  
OTHER INFORMATION: identity 98  
OTHER INFORMATION: region 1..421  
OTHER INFORMATION: id AA280803  
FEATURE:  
NAME/KEY: est  
LOCATION: complement(901..1267)  
IDENTIFICATION METHOD: blastn2  
OTHER INFORMATION: identity 98  
OTHER INFORMATION: region 5..372  
OTHER INFORMATION: id AA724197  
FEATURE:  
NAME/KEY: est  
LOCATION: complement(978..1266)  
IDENTIFICATION METHOD: blastn2  
OTHER INFORMATION: identity 98  
OTHER INFORMATION: region 1..289  
OTHER INFORMATION: id N34970  
FEATURE:  
NAME/KEY: est  
LOCATION: complement(985..1271)  
IDENTIFICATION METHOD: blastn2  
OTHER INFORMATION: identity 98  
OTHER INFORMATION: region 18..304  
OTHER INFORMATION: id AA883384  
FEATURE:  
NAME/KEY: est  
LOCATION: complement(1078..1267)  
IDENTIFICATION METHOD: blastn2  
OTHER INFORMATION: identity 98





```
QY 433 caggagaatgctctcgaacacctagcgataaggcctttgacccggtccatctcgcaggc 492
    |||
Db 218 caggagaatgctctcgaacacctagcgataaggcctttgacccggtccatctcgcaggc 277
    |||
QY 493 cctcctgacacagaggttcttcaatggcattggcacaactatctcgggcagagatcctcgtta 552
    |||
Db 278 cctcctgacacagaggttcttcaatggcattggcacaactatctcgggcagagatcctcgtta 337
    |||
QY 553 ccggtcgaagatccccctttgagaagccgctcgtcgttgcctggagccctgcagcagca 612
    |||
Db 338 ccggtcgaagatccccctttgagaagccgctcgtcgttgcctggagccctgcagcagca 397
    |||
QY 613 caggccagccgagctgacctgagcccaagaagataaagaccacaagctgcagaaatccaga 672
    |||
Db 398 caggccagccgagctgacctgagcccaagaagataaagaccacaagctgcagaaatccaga 457
    |||
QY 673 cctgctgagactatgtcaactcagtgcccaaggaagtgtccagttgggggagagagccta 732
    |||
Db 458 cctgctgagactatgtcaactcagtgcccaaggaagtgtccagttgggggagagagccta 517
    |||
QY 733 cgggtcagagagcgggagagagactttgctcctttcgagccctggctgcgctgctatgg 792
    |||
Db 518 cgggtcagagagcgggagagagactttgctcctttcgagccctggctgcgctgctatgg 577
    |||
QY 793 catgccagcatagctccctcagcagccgcatggcgtaccatctggttccaggggga 852
    |||
Db 578 catgccagcatagctccctcagcagccgcatggcgtaccatctggttccaggggga 637
    |||
QY 853 tctctggaccgttggcaccacaaagggcgcaagtcctccgcaaaagaaatcccaaggccacaca 912
    |||
Db 638 tctctggaccgttggcaccacaaagggcgcaagtcctccgcaaaagaaatcccaaggccacaca 697
    |||
QY 913 gctgagctcctgagacagagcttcttaagagcgtcttgctcccgagcaagcccttccaggac 972
    |||
Db 698 gctgagctcctgagacagagcttcttaagagcgtcttgctcccgagcaagcccttccaggac 757
    |||
QY 973 acgaagggcacaagagagaccttcttaagagcgtcgaaccagcgcctgaggggaccag 1032
    |||
Db 758 acgaagggcacaagagagaccttcttaagagcgtcgaaccagcgcctgaggggaccag 817
    |||
QY 1033 cctccagcaggaccagaagctcccaagtcgccaagta 1070
    |||
Db 818 cctccagcaggaccagaagctcccaagtcgccaagta 855
    |||
RESULT 9
US-60-172-377-10510
; Sequence 10510, Application US/60172377
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; FILE REFERENCE: GX-0008 P
; CURRENT APPLICATION NUMBER: US/60/172,377
; NUMBER OF SEQ ID NOS: 13,749
; SOFTWARE: PERL Program
; SEQ ID NO 10510
; LENGTH: 899
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 903209.3
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 4, 10, 794, 856, 863, 868
; OTHER INFORMATION: a, t, c, g, or other
US-60-172-377-10510
```

```
Query Match 38.3%; Score 536; DB 93; Length 899;
Best Local Similarity 99.7%; Pred. No. 4.8e-242;
Matches 636; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 433 caggagaatgctctcgaacacctagcgataaggcctttgacccggtccatctcgcaggc 492
    |||
Db 218 caggagaatgctctcgaacacctagcgataaggcctttgacccggtccatctcgcaggc 277
    |||
QY 493 cctcctgacacagaggttcttcaatggcattggcacaactatctcgggcagagatcctcgtta 552
    |||
Db 278 cctcctgacacagaggttcttcaatggcattggcacaactatctcgggcagagatcctcgtta 337
    |||
QY 553 ccggtcgaagatccccctttgagaagccgctcgtcgttgcctggagccctgcagcagca 612
    |||
Db 338 ccggtcgaagatccccctttgagaagccgctcgtcgttgcctggagccctgcagcagca 397
    |||
QY 613 caggccagccgagctgacctgagcccaagaagataaagaccacaagctgcagaaatccaga 672
    |||
Db 398 caggccagccgagctgacctgagcccaagaagataaagaccacaagctgcagaaatccaga 457
    |||
QY 673 cctgctgagactatgtcaactcagtgcccaaggaagtgtccagttgggggagagagccta 732
    |||
Db 458 cctgctgagactatgtcaactcagtgcccaaggaagtgtccagttgggggagagagccta 517
    |||
QY 733 cgggtcagagagcgggagagagactttgctcctttcgagccctggctgcgctgctatgg 792
    |||
Db 518 cgggtcagagagcgggagagagactttgctcctttcgagccctggctgcgctgctatgg 577
    |||
QY 793 catgccagcatagctccctcagcagccgcatggcgtaccatctggttccaggggga 852
    |||
Db 578 catgccagcatagctccctcagcagccgcatggcgtaccatctggttccaggggga 637
    |||
QY 853 tctctggaccgttggcaccacaaagggcgcaagtcctccgcaaaagaaatcccaaggccacaca 912
    |||
Db 638 tctctggaccgttggcaccacaaagggcgcaagtcctccgcaaaagaaatcccaaggccacaca 697
    |||
QY 913 gctgagctcctgagacagagcttcttaagagcgtcttgctcccgagcaagcccttccaggac 972
    |||
Db 698 gctgagctcctgagacagagcttcttaagagcgtcttgctcccgagcaagcccttccaggac 757
    |||
QY 973 acgaagggcacaagagagaccttcttaagagcgtcgaaccagcgcctgaggggaccag 1032
    |||
Db 758 acgaagggcacaagagagaccttcttaagagcgtcgaaccagcgcctgaggggaccag 817
    |||
QY 1033 cctccagcaggaccagaagctcccaagtcgccaagta 1070
    |||
Db 818 cctccagcaggaccagaagctcccaagtcgccaagta 855
    |||
```

```
RESULT 10
US-60-172-377-10510
; Sequence 10510, Application US/60172377
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; FILE REFERENCE: GX-0008 P
; CURRENT APPLICATION NUMBER: US/60/172,377
; NUMBER OF SEQ ID NOS: 13,749
; SOFTWARE: PERL Program
; SEQ ID NO 10510
; LENGTH: 899
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 903209.3
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 4, 10, 794, 856, 863, 868
```

OTHER INFORMATION: a, t, c, g, or other  
US-60-172-377-10510

Query Match 38.3%; Score 536; DB 93; Length 899;  
Best Local Similarity 99.7%; Pred. No. 4.8e-242;  
Matches 636; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 433 cagggaagatgtctcagaaacctagcgataaggcctttgacggcccatctgcagagc 492  
DB 218 cagggaagatgtctcagaaacctagcgataaggcctttgacggcccatctgcagagc 277  
QY 493 cctcctggagcagaggtttcaatggcatgtgcaactatctgcggcgagagatcctga 552  
DB 278 cctcctggagcagaggtttcaatggcatgtgcaactatctgcggcgagagatcctga 337  
QY 553 cgggtgaagatccccctttgagaaggcccgctggtcttcggagggccctgcagcaga 612  
DB 338 cgggtgaagatccccctttgagaaggcccgctggtcttcggagggccctgcagcaga 397  
QY 613 caggccagccggagctgacctgagccagagataaggacccaagctgcagaatccaga 672  
DB 398 caggccagccggagctgacctgagccagagataaggacccaagctgcagaatccaga 457  
QY 673 cctgctggagctatgtcactcagtgccccaggaagtgtccagttggggcgagagctga 732  
DB 458 cctgctggagctatgtcactcagtgccccaggaagtgtccagttggggcgagagctga 517  
QY 733 cgggtcagagagcgggagagagactttgctgctttcagcctgctgctgctgctatgg 792  
DB 518 cgggtcagagagcgggagagagactttgctgctttcagcctgctgctgctgctatgg 577  
QY 793 catgcagcagatgctcctcagagcagcggatgcccagctgacctgctgctgctgctatgg 852  
DB 578 catgcagcagatgctcctcagagcagcggatgcccagctgacctgctgctgctgctatgg 637  
QY 853 tcttgaccgttggcaccacaaaggcgcaagtcccgcaaaagaaatccaaagccacaga 912  
DB 638 tcttgaccgttggcaccacaaaggcgcaagtcccgcaaaagaaatccaaagccacaga 697  
QY 913 gctgagctctgagagagagagagctttgctcctcagagcaagggcccttccagggac 972  
DB 698 gctgagctctgagagagagagagagctttgctcctcagagcaagggcccttccagggac 757  
QY 973 acgaagggcaagagagagagagcttctcagagagagctgcaacccagcggcctgaggggaccag 1032  
DB 758 acgaagggcaagagagagagagcttctcagagagagctgcnaccagcggcctgaggggaccag 817  
QY 1033 cctccagcagggaccagagagctcccaagtgcccaaga 1070  
DB 818 cctccagcagggaccagagagctcccaagtgcccaaga 855

RESULT 11

US-09-471-275-1784/c  
Sequence 1784, Application US/09471275  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: Novel Contigs Obtained  
TITLE OF INVENTION: From Various Libraries  
FILE REFERENCE: 782  
CURRENT APPLICATION NUMBER: US/09/471,275  
EARLIER FILING DATE: 1999-12-23  
EARLIER APPLICATION NUMBER: US 09/235,076  
EARLIER FILING DATE: 1999-01-20  
EARLIER APPLICATION NUMBER: US 09/234,611  
EARLIER FILING DATE: 1999-01-22  
EARLIER APPLICATION NUMBER: US 09/240,371  
EARLIER FILING DATE: 1999-01-29  
EARLIER APPLICATION NUMBER: US 09/277,227  
EARLIER FILING DATE: 1999-03-25  
EARLIER APPLICATION NUMBER: US 09/271,490  
EARLIER FILING DATE: 1999-03-18

EARLIER APPLICATION NUMBER: US 09/293,972  
EARLIER FILING DATE: 1999-04-15  
EARLIER APPLICATION NUMBER: US 09/274,861  
EARLIER FILING DATE: 1999-03-23  
EARLIER APPLICATION NUMBER: US 60/125,453  
EARLIER FILING DATE: 1999-03-19  
EARLIER APPLICATION NUMBER: US 60/126,605  
EARLIER FILING DATE: 1999-03-26  
EARLIER APPLICATION NUMBER: US 09/306,350  
EARLIER FILING DATE: 1999-05-07  
EARLIER APPLICATION NUMBER: US 09/399,720  
EARLIER FILING DATE: 1999-09-21  
EARLIER APPLICATION NUMBER: US 09/404,284  
EARLIER FILING DATE: 1999-09-21  
EARLIER APPLICATION NUMBER:  
EARLIER FILING DATE: 1999-12-16  
NUMBER OF SEQ ID NOS: 10451  
SOFTWARE: pt.ct\_genes Version 1.0  
SEQ ID NO 1784  
LENGTH: 1451  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (535)...(116)  
OTHER INFORMATION: Similar to g1552075 in the genepept database release 114,  
OTHER INFORMATION: Run with FASTXY 3.3t00, default parameters  
US-09-471-275-1784

Query Match 32.0%; Score 448; DB 54; Length 1451;  
Best Local Similarity 99.8%; Pred. No. 1.5e-200;  
Matches 498; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 874 agggcgcaagtcctcccaaaaataatcaaggccacacagctgagtcctcgagacagat 933  
DB 530 AGGGCGCAAGTCCCGCAAAAATAATCCAGGCCACACAGCTGAGTCTGAGGACAGAT 471  
QY 934 ggagacgctttgctccgagcaaggcccttccaggacacagaaagggcaagagacat 993  
DB 470 GGAGACGCTTTGCTCCCAAGCAAGGCCCTTCCAGGACACAGAGGACAGACCT 411  
QY 994 tcttaagaggaactgaacccagcggcctgagggagccagcctccagcaggaccagagac 1053  
DB 410 TCTTAAGAGGACTGCAACCCAGCGCCTGAGGGGACCAGCCTCCAGCAGGACCCAGAGC 351  
QY 1054 tcccaagtgcccaagaaaggagagagagagagagagagagagagagagagag 1113  
DB 350 TCCCAAGTGCCCAAGAGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 291  
QY 1114 acccgggaaggtcaaggctgacatcccatccttggaaacagagggagccctccttcta 1173  
DB 290 ACCCGGAAGGTCAAGGCTGACATCCCATCCTTGAACACAGAGGGAGGACCTCACTTA 231  
QY 1174 gcaggagcctcctctgcttgctgacacaccccttcttattgtcttgcctgcactgaggg 1233  
DB 230 GCAGGAGGCTCTCCTTGCTTGCTGACTCAGCCTTTCTTATTGTCTTGCCTGCTGAGGG 171  
QY 1234 tctgaattttgggagcaggaataatctgaagtgcaaacagccctcagcgtgttccct 1293  
DB 170 TCTGAATTTTGGGAGCAGGCANTATCTGAAGGTCAAAACAGGCCCTACGGCTGTTCCT 111  
QY 1294 gcaaacactctcaggttttaattgtaccacctcttccacctctttaaagctcattgaaa 1353  
DB 110 GCACAACCTCTCATGTTTAAATTGATCCCATCTTCCACATCTTTAAAGCTCATGTGAA 51  
QY 1354 atgctgcatttttaataa 1372  
DB 50 AATGCTGCATTTTAAATAA 32

RESULT 12  
US-09-277-227-17958







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

```
Run on: May 1, 2000, 14:44:09 ; Search time 3022.95 Seconds
          (without alignments)
          1746.102 Million cell updates/sec
```

**Title:** US-09-215-435-71  
**Perfect score:** 1398  
**Sequence:** 1 gatgcctgagggcccccagc.....acatttgaaaaaaaaaa 1398

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4538634 seqs, 1887831982 residues

```
Word size : 0
Total number of hits satisfying chosen parameters: 9077268
```

```
Minimum DB seq length: 0
Maximum DB seq length: 1000000
```

Post-processing: Listing first 45 summaries

```
Database : EST.*
1:  em_est1:*
2:  em_est2:*
3:  em_est3:*
4:  em_est4:*
5:  em_est5:*
6:  em_est6:*
7:  em_est7:*
8:  em_est8:*
9:  em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
```

44:	gb_est225:
45:	gb_est226:
46:	gb_est27:
47:	gb_est28:
48:	gb_est29:
49:	gb_est30:
50:	gb_est31:
51:	gb_est32:
52:	em_est20:
53:	em_est21:
54:	em_est22:
55:	em_est23:
56:	em_est24:
57:	em_est25:
58:	em_est26:
59:	gb_est33:
60:	gb_est34:
61:	gb_est35:
62:	gb_est36:
63:	gb_est37:
64:	gb_est38:
65:	em_est27:
66:	em_est28:
67:	em_est29:
68:	em_est30:
69:	em_est31:
70:	gb_est40:
71:	gb_est41:
72:	gb_est42:
73:	gb_est43:
74:	gb_est44:
75:	em_est31:
76:	em_est32:
77:	em_est33:
78:	em_est34:
79:	gb_gss1:
80:	gb_gss2:
81:	gb_gss3:
82:	gb_gss4:
83:	em_gss1:
84:	em_gss2:
85:	em_gss3:
86:	em_gss4:
87:	gb_gss5:
88:	gb_gss6:
89:	gb_gss7:
90:	gb_gss8:
91:	gb_gss9:
92:	em_gss5:
93:	em_gss6:
94:	em_gss7:
95:	em_gss8:
96:	em_gss9:
97:	em_gss10:
98:	em_gss11:
99:	gb_gss10:
100:	gb_gss11:
101:	em_gss12:
102:	gb_gss12:
103:	gb_gss13:
104:	gb_gss14:
105:	gb_gss15:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	433	31.0	450	62	A1925612 w034c09.3

```
c 2 433 31.0 498 62 AI913135 wa10a10.x
c 3 387 27.7 498 45 AI341866 qt10d04.x
c 4 380 27.2 380 49 AI638424 t31g11.x
c 5 380 27.2 401 45 AI393566 tg46c11.x
c 6 378 27.0 403 48 AI608804 tw21e10.x
c 7 377 27.0 377 45 AI373305 q245a03.x
c 8 347 24.8 384 63 AW014988 UI-H-B10p
c 9 329 23.5 329 44 AI290102 qw26c09.s
c 10 329 23.5 521 38 AA744383 ny63d09.s
c 11 317 22.7 419 35 AA595778 nj28h10.s
c 12 284 20.3 384 61 AI965117 wk08a10.x
c 13 280 20.0 478 36 AA604501 no73b09.s
c 14 259 18.5 412 42 AI079757 oy43b02.s
c 15 257 18.4 512 39 AA833865 oc62h07.s
c 16 243 17.4 372 37 AA724197 ai09d12.s
c 17 243 17.4 505 73 AW170234 xn57d05.x
c 18 242 17.3 388 39 AA832090 oc94b09.s
c 19 233 16.7 372 24 N34970 yy50g09.s
c 20 232 16.6 306 37 AA668993 aa81c10.s
c 21 218 15.6 403 61 AI816095 au43c02.y
c 22 217 15.5 421 31 AA280803 zs99a08.s
c 23 216 15.5 216 44 AI280739 qw07c03.x
c 24 210 15.0 504 39 AA831660 oc83g09.s
c 25 196 14.0 497 22 R62755 y110b09.r1
c 26 188 13.4 360 50 AI700271 w03f09.x
c 27 187 13.4 217 63 AI934214 wn97g03.x
c 28 186 13.3 430 22 R63100 y101h08.s1
c 29 178 12.7 480 61 AI816136 au45c02.x
c 30 176 12.6 396 39 AA824509 oc82e09.s
c 31 176 12.6 461 22 R62756 y110b09.s1
c 32 168 12.0 304 39 AA883584 al46c05.s
c 33 152 10.9 313 46 AI458543 tk04a04.x
c 34 136 9.7 401 24 H78102 yu83f11.s1
c 35 128 9.2 128 37 AA731925 nz83b02.s
c 36 121 8.7 203 40 AA927220 om17a10.s
c 37 108 7.7 186 26 W80730 zh51e02.s1
c 38 89 6.4 331 24 H65509 yu52e11.s1
c 39 70 5.0 390 38 AA739446 vv53f01.r
c 40 67 4.8 143 32 AA359490 EST68516
c 41 65 4.6 325 26 W44504 zc28a12.r1
c 42 53 3.8 269 25 N44044 yy50g09.r1
c 43 42 3.0 456 26 W78831 zh51e02.r1
c 44 41 2.9 376 61 AI867438 wb71d03.x
c 45 41 2.9 446 49 AI656745 tt53n06.x
```

## ALIGNMENTS

```
RESULT 1
AI925612/c 450 bp mRNA EST 02-SEP-1999
LOCUS wa34c09.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2457232 3',
DEFINITION mRNA sequence.
ACCESSION AI925612
VERSION AI925612.1 GI:5661663
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 450)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncilogap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On May 18, 1998 this sequence version replaced gi:3137298.
Contact: Robert Strausberg, Ph.D.
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
```

```
cdNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 406.
Location/Qualifiers
Source
1. 450
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2457232"
/clone_lib="NCI_CGAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signed ring cell features"
/lab_host="DHL08"
/Note="Organ: Stomach; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"

BASE COUNT 96 a 114 c 129 g 111 t
ORIGIN

Query Match 31.0%; Score 433; DB 62; Length 450;
Best Local Similarity 100.0%; Pred. No. 1.5e-195;
Matches 433; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 953 agcaagggccctccaggacagagggcaagagagaccttccttaaggagactgaacc 1012
|||||
Db 433 AGCAAGGCCCTTCCAGGACACAGAGGGCAAGAGAGAGACTTCTTAAGAGGACTGCAACC 374
|||||

QY 1013 cagcgccctgaggggaccagcctccagcaggaccagagctcccccagtgcccaagaag 1072
|||||
Db 373 CACGCGCCTGAGGGGACCAGCCTCCAGCAGGACCCAGAGAGCTCCACAGTGCCCAAGAG 314
|||||

QY 1073 gggagaggaagggggcagcaggcagcctctggccactgcagaccccggaaggtcaaggct 1132
|||||
Db 313 GGGAGGAGGAGGGGGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 254
|||||

QY 1133 gacatcccatccttgaaacagaggggacactcagcctcttagcaggaggtctcttgcct 1192
|||||
Db 253 GACATCCCATCCTTGAACACAGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 194
|||||

QY 1193 tgcactcacccttcttattgtcttgcctgcactctgggggtctgaatttttgggagcag 1252
|||||
Db 193 TGCACACACCTTCTTATGTCTTGCCTGCATCTGGGGGTCTGAATTTTGGGAGCAG 134
|||||

QY 1253 gcaatctcgaaggtgcaaacaggccctcagcgtgttcctcgcacactctcattgtttt 1312
|||||
Db 133 GCAATATCTGAAGGTGCAACAGGCGCTACGGCTGTTCCTGCACAACTCTCATGTTTT 74
|||||

QY 1313 aattgaccccatctccacatctttaagctcatgtgaaaatgctgcatcttttaataa 1372
|||||
Db 73 AATTGACCCCATCTTCCACATCTTTAAAGCTCATGTGAAAAATGCTGCAATTTTAAATAA 14
|||||

QY 1373 actgatacatttg 1385
|||||
Db 13 ACTGATACATTG 1

RESULT 2
AI913135/c 498 bp mRNA EST 28-JUL-1999
LOCUS wa10a10.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2297658 3',
DEFINITION mRNA sequence.
ACCESSION AI913135
VERSION AI913135.1 GI:5632990
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```









Fri May 12 12:27:59 2000

```

REFERENCE 1 (bases 1 to 377)
AUTHORS  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Jun 15, 1998 this sequence version replaced gi:3223423.
          Contact: Robert Strausberg, Ph.D.
          Tel: (301) 496-1550
          Email: Robert.Strausberg@nih.gov
          Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
          Emmert-Buck, M.D., Ph.D.
          cDNA Library Prepared by: M. Bento Soares, Ph.D.
          cDNA Library Arrayed by: Greg Lennon, Ph.D.
          DNA Sequencing by: Washington University Genome Sequencing Center
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 516      Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 371.
          Location/Qualifiers
FEATURES             source
1..377
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_Kid1"
/lab_host="DH10B"
/note="Organ: kidney; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI-CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT      86 a 94 c 107 g 90 t
ORIGIN
Query Match      27.0%; Score 377; DB 45; Length 377;
Best Local Similarity 100.0%; Pred. No. 6.6e-169;
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1010 acccagcgctgagggaccagctccagcaggaccagaccagctccacagtgcccaag 1069
|||||
Db 377 ACCCAGCGCCTGAGGGACCAGCCTCCAGGAGCCAGGACCCAGGACCTCCAGTCCCAAG 318
QY 1070 aaggagagagagggcgacgagcagcctctggcactcagaccgccggagaggtcaag 1129
|||||
Db 317 AAGGGAGAGAGAGGGCGACGACGACCCCTTGCCACTCGAGACCCCGGAGGTCAAG 258
|||||
QY 1130 gctgacatcccatctgttgaaacagaggggagcctcagcctcttgcagagaggtctcctt 1189
|||||
Db 257 GCTGACATCCCATCTTGTGAACAGAGGGGAGCCTCAGCCTCTTAGCAGGAGGCTCCTT 198
QY 1190 gcttgacatccatcttattgtcttgcctgcctgcctgcctgcctgcctgcctgcctgcct 1249
|||||
Db 197 GCTTGACATCCATCTTGTATGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 138
QY 1250 caggcaatctgaaggtgcaaacaggccctcagcctgttccctgcacaaactcctcaggt 1309
|||||
Db 137 CAGGCAATCTGAAGTGCACAAAGGCCCTACGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 78
QY 1310 ttaattgtacccatcttcccatctttaaagctcatgtgaaatctgcatttttaa 1369
|||||
Db 77 TTTAATTGTACCCCATCTTCCATCTTTAAAGCTCATGTGAAATGCTGCTGCTGCTGCTGCT 18
QY 1370 taaactgatacatttga 1386

```

```

|||||
Db 17 TAACTGATACATTGA 1
|||||
RESULT 8
LOCUS  AW014988/c
DEFINITION  UI-H-BIOP-aaz-d-07-0-UI-s1 NCI-CGAP_Sub2 Homo sapiens cDNA clone
          IMAGE:2711077 3', mRNA sequence.
ACCESSION  AW014988
VERSION  AW014988.1 GI:5863745
KEYWORDS  EST.
SOURCE  human.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 384)
          NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On May 18, 1998 this sequence version replaced gi:3138294.
          Contact: Robert Strausberg, Ph.D.
          Tel: (301) 496-1550
          Email: Robert.Strausberg@nih.gov
          The sequence contained an oligo-dT track that was present in the
          oligonucleotide that was used to prime the synthesis of first
          strand cDNA and therefore this may represent a bonafide poly A
          tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
          NCI-CGAP clone distribution information can be found through the
          I.M.A.G.E. Consortium/LLNL at:
          www-bio.llnl.gov/bbrp/image/image.html
          Seq primer: M13 Forward
          POLYA=Yes.
          Location/Qualifiers
FEATURES             source
1..384
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_Sub2"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NCI-CGAP_Sub2 library is a subtracted library derived from
BI-BI constitutes a mixture of 21 normalized or
subtracted NCI-CGAP libraries: NCI-CGAP_Co4,
NCI-CGAP_Pr22, NCI-CGAP_Pr28, NCI-CGAP_Co10,
NCI-CGAP_Co16, NCI-CGAP_Kid3, NCI-CGAP_Kid5, NCI-CGAP_Kid12,
NCI-CGAP_Br2, NCI-CGAP_Co8, NCI-CGAP_Co11, NCI-CGAP_Lu24,
NCI-CGAP_Br23, NCI-CGAP_Lu5, NCI-CGAP_GC6, NCI-CGAP_Le12,
NCI-CGAP_Lu19, NCI-CGAP_GC4, NCI-CGAP_GC6, NCI-CGAP_Br25.
These 21 libraries were pooled and a single-stranded DNA
preparation of the resulting mixture was used as a tracer
in a subtractive hybridization with a driver whose
composition is detailed below: NCI-CGAP_Kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE ClonesIDs
132376-1323911, 1456008-1456775, 1500552-1502855)
NCI-CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
(IMAGE ClonesIDs 1323912-1325831, 1471368-1472903,
1492104-1493255) NCI-CGAP_Lu5 pool 1 LLAM 3575-3582,
3851-3854 (IMAGE ClonesIDs 1414920-1417991,
1520904-1522439) NCI-CGAP_GC4 pool 1 LLAM 3164-3167,
3716-3720, 3733-3735 (IMAGE ClonesIDs 1257096-1258631,
1469064-1470983, 1475592-1476743) NCI-CGAP_Pr22 pool 1
LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE ClonesIDs
985608-986759, 1101192-1101959, 1217928-1220615)
NCI-CGAP_Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE
ClonesIDs 1057416-1061255, 1144584-1145351) The resulting
subtracted library contained 4 million recombinants.
Subtraction was performed as previously described
[Bonaldo, Lennon & Soares (1996): Normalization and
Subtraction: Two Approaches To Facilitate Gene Discovery.

```



```

Insert Length: 770 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 488.

FEATURES
    source
        1..521
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="NCI_CGAP_GCB1"
            /tissue_type="germinal center B cell"
            /lab_host="DH10B"
            /note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD+),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-GTGTACCAATCTGAAGTGGAGCGCCGCTCATTTTTTTTTTTTTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      103 a   132 c   143 g   143 t
ORIGIN
Query Match      23.5%; Score 329; DB 38; Length 521;
Best Local Similarity 99.5%; Pred. No. 4.6e-146;
Matches 429; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 966 ccaggacacgaaggcaagagagaccttcttaagagagacttcaaccagcgccctgag 1025
Db 431 CCAGGACACGAGGCAAGAGAGAGACCTTCTTAAGAGAGACTGCAACCCAGCGCCTGAGG 372
Qy 1026 ggaccagcctcagcagagcccaagaagctccaccagtgcccaagaaggaggaggaagg 1085
Db 371 GGACCAAGCCTCCAGCAGGACCCAGAAAGCTCCACAGTGCCCAAGAGAGGAGGAGAGG 312
Qy 1086 ggcacagggagcctctggccactgcagaccccggaagggtcaaggctgacatccatcct 1145
Db 311 GGCACAGGAGGAGCCTCTGGCCACTGCAGACCCCGGAAGGTCAAGGCTGACATCCCATCT 252
Qy 1146 tggaaccagaggggacatcagcctcttagcaggagggtctccttgcctgcactacacctt 1205
Db 251 TGGAAACAGAGGGGACCTCAGCCTCTTAGCAGGAGGCTCTCCTTGTCTGCACCTCACCTT 192
Qy 1206 tcttatgtcttgcctgcactctggggtctgaatttttgggagcagcgaatctgaag 1265
Db 191 TCTTATTGTCTTGCCCTGCATCTGGGGGTCTGAATTTTGGGAGCAGGCAATATCTGAAG 132
Qy 1266 gtgcaaacagggcctcagggctgttccctgcacaactctcattggttttaattgtaccoccat 1325
Db 131 GTGCACATGAGGCTCAGGCTGTTCCCTGCACACTCTCATGGTTTAATGTATCCCAT 72
Qy 1326 ctccacatctttaaagctcatgtgaaaaatgctgcattttttaataaactgatacatgtg 1385
Db 71 CTTCCATCTCTTAAAGCTCATGTGAAAAATGCTGCATCTTTTAAATAAAGTATGATATTG 12
Qy 1386 aaaaaa 1396
Db 11 AAAAAAAAAA 1

RESULT 11
LOCUS AA595778 419 bp mRNA EST 25-SEP-1997
DEFINITION nJ28h10.s1 NCI_CGAP_A41 Homo sapiens cDNA clone IMAGE:593859 3',
mRNA sequence.
ACCESSION AA595778
VERSION AA595778.1 GI:2411128
KEYWORDS EST.

```

```

SOURCE      Homo sapiens
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 419)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:693024.
            Contact: Robert Strausberg, Ph.D.
            Tel.: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmerit-Buck, M.D., Ph.D.
            cDNA Library Preparation: Stratagene, Inc., David B. Krizman,
            Ph.D.
            cDNA Library Arraying: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 871 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 419.
Location/Qualifiers
    1..419
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_lib="IMAGE:993859"
        /clone_host="NCI_CGAP_A41"
        /tissue_type="adrenal adenoma"
        /lab_host="SOLR (kanamycin resistant)"
        /note="Organ: adrenal gland; Vector: Bluescript SK-;
            Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally.
            Primer: Oligo dT. Two pooled bulk adrenal adenomas. 5'
            adaptor sequence: 5' GAATTCGGCACGAG 3' 3' adaptor
            sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' Average insert
            size: 1.6 kb."
BASE COUNT      90 a   106 c   118 g   105 t
ORIGIN
Query Match      22.7%; Score 317; DB 35; Length 419;
Best Local Similarity 99.5%; Pred. No. 2.4e-140;
Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 967 caggacacgaaggcaagagagaccttcttaagagagacttcaaccagcgccctgaggg 1026
Db 419 CAGGACACGAAGGGCAAGAGAGACCTTCTTAAGAGAGACTGCAACCCAGCGCCTGAGGG 360
Qy 1027 gaccagcctccagcagggagcccaagaagctcccaagtgcccaagaaggaggaggaagg 1086
Db 359 GACCACGCTCCAGCAGGACCCAGAGAGTCCACAGTGCCCAAGAGGGGAGGAGGAGG 300
Qy 1087 ggaacaggcagcctctgcccactgcagaccccggaagggtcaaggctgacatccatcctt 1146
Db 299 GCACAGGCGCCTCTTGCCACTGCAGACCCCGGAAGGTCAAGGCTGACATCCCATCTCT 240
Qy 1147 ggaacacaggggagcctcagcctcttagcaggaggtctccttgcctgcactcacacctt 1206
Db 239 GGAACCCAGAGGGACCTCAGCCTCTTAGCAGGAGGCTCTCCTTGTCTGCACCTCACCTT 180
Qy 1207 cttattgtcttgcctgcactctgggggtctgaatttttgggagcagggcaaatatctgaagg 1266
Db 179 CTTATTGTCTTGCCCTGCATCTGGGGGTCTGAATTTTGGGAGCAGGCAATATCTGAAGG 120
Qy 1267 tgaacaacagggcctcagcctgttccctgcacaactcctatgttttaattgtaccccatc 1326
Db 119 TGCAAAACAGGCCCTACGGGTGTCCCTGCACAACCTCTCATGTGTTTAAATTGATCCCATC 60
Qy 1327 ttccacatctttaaagctcatgtgaaaaatgctgcattttttaataaactgatacatgtg 1385

```

```
Db 59 TCCACATCTTTAAAGCTCATGTGAAATGCTGATTTTAAATAAAGTACATTTG 1
|||||
RESULT 12
AI865117/c 384 bp mRNA EST 30-AUG-1999
LOCUS wk08a10.x1 NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2411706 3',
DEFINITION mRNA sequence.
ACCESSION AI865117
VERSION AI865117
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 384)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1998 this sequence version replaced gi:3137929.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Life Technologies catalog #: 11547-015
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 375.
Location/Qualifiers
1. 384
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2411706"
/tissue_type="lymphoma, follicular mixed small and large
cell"
/lab_host="DH10B"
/notes="Organ: lymph node; Vector: pCMV-SPORT6; Site_1:
SalI; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dt. Average insert size 1.25 kb. Life Technologies
catalog #: 11547-015"
BASE COUNT 102 a 94 c 103 g 85 t
ORIGIN
Query Match 20.3%; Score 284; DB 61; Length 384;
Best Local Similarity 100.0%; Pred. No. 1.2e-124;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1103 gcccctcagaccggaggtcaaggtgacatccatcccttggaaacagagggacc 1162
Db 284 GCCACTGCAGACCCCGGAGGTCAAGGCTGACATCCCATCCCTTGGAAACAGAGGGGACC 225
QY 1163 tcagcctcttagcaggaggtctccttgcctgactcacccttcttattgtcttgccct 1222
Db 224 TCAGCCTCTTAGCAGGAGGCTCTCCTTGTGTCATCACCCTTCTTATTGTCTTGCCCT 165
QY 1223 gcatctgggggtctgaatttttggagcaggccaatatctgaaggtygcaaacaggccctac 1282
Db 164 GCATCTGGGGGTCTGAATTTTGGGAGCAGGCAATATCTGAAGTGTCAAAACAGGCCCTAC 105
QY 1283 ggcgttccctgcacaaactctcatggttttaattgtaccccattcttccacatctttaag 1342
Db 104 GGCCTGTCCCTGCACAACTCTCATGTGTTTAAATGTGTACCCCATCTTCCACATCTTTAAAG 45
QY 1343 ctcatgtgaaaatgctgatattttaataaactgatacatattga 1386
Db 44 CTCATGTGAAAATGCTGCATTTTNTAANAAGTACATTTGA 1
```

```
RESULT 13
AA604501/c 478 bp mRNA EST 08-OCT-1997
LOCUS no73b09.s1 NCI_CGAP_AAL Homo sapiens cDNA clone IMAGE:1112441 3',
DEFINITION mRNA sequence.
ACCESSION AA604501
VERSION AA604501
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 478)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1406949.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc., David B. Krizman,
Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 879 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 471.
Location/Qualifiers
1. 478
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1112441"
/clone_lib="NCI_CGAP_AAL"
/tissue_type="adrenal adenoma"
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: adrenal gland; Vector: Bluescript SK-;
Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally.
Primer: Oligo dt. Two pooled bulk adrenal adenomas. 5'
adaptor sequence: 5' GAATTCGGCAGCAG 3' 3' adaptor
sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3' Average insert
size: 1.6 kb.
BASE COUNT 100 a 125 c 134 g 119 t
ORIGIN
Query Match 20.0%; Score 280; DB 36; Length 478;
Best Local Similarity 99.3%; Pred. No. 9.6e-123;
Matches 430; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 953 agcaagggcccttcagacacagagggcaagagagaccttccaaagagactcaacc 1012
Db 433 AGCAAGGGCCCTTCCAGGACAGAGGCAAGAGAGAGACTTCTTAAAGAGACTGCAACC 374
QY 1013 cagcggccttaggggaccagcctccagcagagaccagagctccacagtggtcccaagaag 1072
Db 373 CAGCGGCTTAGGGGACAGCCTCCAGCAGGACCCAGAAAGCTCCACAGTGCCTCAAGAG 314
QY 1073 gggagaggaagggggagacagagcctcttggccactgcagaccgggaaggtcaagpct 1132
Db 313 GGGAGGAGAGAGGGGCGCAGACGCTCTGGCCACTGTGCCACTCGAGACCCCGGAAGGCT 254
QY 1133 gacatcccatcttggaaaccagaggggaccttagcagcctttagcagaggtctcttct 1192
Db 253 GACATCCCATCTTGGAAACAGAGGGGACCTCAGCCTCTTAGCAGGAGGCTCTCCTTGT 194
```

QY 1193 tgcactcacccttcttattgtctgcctgcactcgtggggtctgaatttttgggagcag 1252  
 Db 193 TGCACACACCTTCTATTGTCTTGCCTGCATCTGGGGTCTGAATTTGGGAGCAG 134  
 QY 1253 gcaatctgaaggtgcaaacagccctacggtgtctcctgcacactctcatgtttt 1312  
 Db 133 GCAATATCTGAAGGTGCAACAGGGGCTACGGGTGTTCCTGTCACAACTCTCATGCTTT 74  
 QY 1313 aattgacccactctccacacttttaagctcatgtgaaaaatgctgatttttaataa 1372  
 Db 73 AATTGACCCCACTTCCACATCTTTAAAGCTCATGTGAAAAATGCTGATTTTAAATA 14  
 QY 1373 actgatacatgtt 1385  
 Db 13 ACTGATACATTG 1

RESULT 14  
 AI079757/c  
 LOCUS  
 DEFINITION OY43b02.s1 Soares\_parathyroid\_tumor\_NbHPA Homo sapiens cDNA clone IMAGE:1668555 3', mRNA sequence.  
 ACCESSION AI079757  
 VERSION AI079757.1 GI:3416008  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 412)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On Apr 18, 1995 this sequence version replaced gi:775324.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 378.  
 Location/Qualifiers  
 1. .412  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1668555"  
 /clone\_lib="Soares\_parathyroid\_tumor\_NbHPA"  
 /tissue\_type="parathyroid tumor"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: parathyroid gland; Vector: pT73D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer  
 [5'  
 TGTACCAATCTGAAGTGGAGGGCGGCACCAATTTTTTTTTTTTTTTTTTTTTT  
 T-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."

BASE COUNT 90 a 102 c 112 g 108 t

ORIGIN

Query Match 18.5%; Score 259; DB 42; Length 412;  
 Best Local Similarity 99.3%; Pred. No. 9.5e-113;  
 Matches 409; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 984 agagagaccttcttaagaggactcaaccagcgccctgaggggaccagcctccagcagg 1043  
 Db 412 AGAGAGACCTTCTTAAGAGGACTGCAACCGCGCTGAGGGGACGCTCCAGCAGG 353  
 QY 1044 acccagaagctcccaagcagtgcccaagaaggaggaggaaggagggcgagcagcctctg 1103  
 Db 352 ACCCAGAAGCTCCACAGTGCCCAAGAGAGGGGAGGAGGAGGACAGGCACGCTCTG 293  
 QY 1104 gccactgagaccccggaaggtcaaggctgacatcccatccttggaaaccagagggaacct 1163  
 Db 292 GCCACTGCAGACCCCGGAAGGTCAAGGCTGACATCCCATCTCTTGAACACAGAGGGACCT 233  
 QY 1164 cagctcttagcagagggctctccttctgttgcactcaacccttcttattgtcttgcctg 1223  
 Db 232 CAGCCTCTTAGCAGAGGCTCTCCTTGTCTGCTGACCTCACCTTCTTATTGTCTTGCCTG 173  
 QY 1224 catctgggggtctgaatttttgggagcgaggaatctctgaagggtgcaaacagggcctctag 1283  
 Db 172 CATCTGGGGGTCTGAATTTTGGGAGCAGGCAATATCTGAAGGTGCAACACAGCGCTACG 113  
 QY 1284 gctgttccctgcacactctcattgttttaattgtaccccatcttccacatctttaaagc 1343  
 Db 112 GCTGTTCCTGCACAACTCTCATGTTTAAATGTACCCCATCTCCACATCTTTAAAGC 53  
 QY 1344 tcattgaaaaatgctgcatcttttaataaacatgatacatattgaaaaaaaaa 1395  
 Db 52 TCATGTGAAATGCTGCATTTTAAATAACTGATACATTTGAAAAAATAA 1

RESULT 15  
 AA833865/c  
 LOCUS  
 DEFINITION od62h07.s1 NCI\_CGAP\_GCBI Homo sapiens cDNA clone IMAGE:1372573, mRNA sequence.  
 ACCESSION AA833865  
 VERSION AA833865.1 GI:2908633  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 512)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On May 8, 1995 this sequence version replaced gi:800233.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 646 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 485.  
 Location/Qualifiers  
 1. .512

FEATURES  
 source





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:16:43 ; Search time 4425.31 Seconds  
(without alignments)  
-360.910 Million cell updates/sec

Title: US-09-215-435-76

Perfect score: 526

Sequence: 1 ctgctgtgctgtgtgcac.....cttctgtcaaaaaaaaaa 526

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 821193 seqs, -1518192014 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_pal:\*

2: gb\_ba2:\*

3: gb\_om:\*

4: gb\_ov:\*

5: gb\_pat:\*

6: gb\_ph:\*

7: gb\_p12:\*

8: gb\_p12:\*

9: gb\_p12:\*

10: gb\_p12:\*

11: gb\_p12:\*

12: gb\_p12:\*

13: gb\_p12:\*

14: gb\_p12:\*

15: gb\_p12:\*

16: gb\_p12:\*

17: gb\_p12:\*

18: gb\_p12:\*

19: gb\_p12:\*

20: gb\_p12:\*

21: gb\_p12:\*

22: gb\_p12:\*

23: gb\_p12:\*

24: gb\_p12:\*

25: gb\_p12:\*

26: gb\_p12:\*

27: gb\_p12:\*

28: gb\_p12:\*

29: gb\_p12:\*

30: gb\_p12:\*

31: gb\_p12:\*

32: gb\_p12:\*

33: gb\_p12:\*

34: gb\_p12:\*

35: gb\_p12:\*

36: gb\_p12:\*

37: gb\_p12:\*

38: gb\_p12:\*

39: gb\_p12:\*

40: gb\_p12:\*

41: gb\_p12:\*

42: gb\_p12:\*

43: gb\_p12:\*

44: gb\_htg6:\*

45: gb\_htg7:\*

46: em\_htg1:\*

47: em\_htg2:\*

48: em\_htg3:\*

49: em\_hum5:\*

50: gb\_pl3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	202	38.4	33458	11	HSDJ60101	AL109656 Human DNA
c 2	20	3.8	107436	8	ATAC006068	AC006068 Arabidops
c 3	20	3.8	152143	33	AC004932	AC004932 Homo sapi
c 4	20	3.8	173338	41	AC009560	AC009560 Homo sapi
c 5	19	3.6	1027	8	AF061577	AF061577 Oryza sat
c 6	19	3.6	2165	12	MMU96685	U96685 Mus musculu
c 7	19	3.6	10808	2	AE001235	AE001235 Treponema
c 8	19	3.6	19839	40	AF094481	AF094481 Homo sapi
c 9	19	3.6	34379	9	HUMCC1S149	MG3005 Homo sapien
c 10	19	3.6	38702	33	AC006646	AC006646 Caenorhab
c 11	19	3.6	40873	32	CEH04109	292848 Caenorhabdi
c 12	19	3.6	41369	10	CH19F15314	AD000091 Homo sapi
c 13	19	3.6	42141	34	CEY54G9A	AL032648 Caenorhab
c 14	19	3.6	52597	11	HSJ421120	AL117327 Human DNA
c 15	19	3.6	58864	9	HUMHDABCD	MG3544 Human DNA s
c 16	19	3.6	89035	8	ATT22F8	AL050351 Arabidops
c 17	19	3.6	91863	8	ATAC004667	AC004667 Arabidops
c 18	19	3.6	94205	40	AC007639	AC007639 Homo sapi
c 19	19	3.6	105914	41	AC008486	AC008486 Homo sapi
c 20	19	3.6	110000	32	CEY53C10_3	Continuation (4 of
c 21	19	3.6	118610	33	AC002344	AC002344 Homo sapi
c 22	19	3.6	127594	41	AC008630	AC008630 Homo sapi
c 23	19	3.6	128269	41	AC011338	AC011338 Homo sapi
c 24	19	3.6	131603	45	AC016962	AC016962 Homo sapi
c 25	19	3.6	146058	40	AC003668	AC003668 Homo sapi
c 26	19	3.6	152826	32	AP000835	AP000835 Homo sapi
c 27	19	3.6	155785	42	AC012109	AC012109 Homo sapi
c 28	19	3.6	165445	42	AC009969	AC009969 Homo sapi
c 29	19	3.6	169763	42	AC011042	AC011042 Homo sapi
c 30	19	3.6	172984	11	HUU91318	U91318 Human chrom
c 31	19	3.6	176836	44	AC016773	AC016773 Homo sapi
c 32	19	3.6	179905	42	AC011751	AC011751 Homo sapi
c 33	19	3.6	261437	41	AC011492	AC011492 Homo sapi
c 34	19	3.6	296589	32	CEY47H10	295311 Caenorhabdi
c 35	19	3.6	300197	32	CEY54G9	298869 Caenorhabdi
c 36	18	3.4	56	9	HUMUT679A	L31744 Human STS U
c 37	18	3.4	320	13	G27350	G27350 human STS S
c 38	18	3.4	336	35	AF127464	AF127464 Mytilus s
c 39	18	3.4	336	35	AF127465	AF127465 Mytilus s
c 40	18	3.4	336	35	AF127466	AF127466 Mytilus s
c 41	18	3.4	336	35	AF127467	AF127467 Mytilus s
c 42	18	3.4	467	13	HUMUT6360	L30549 Human STS U
c 43	18	3.4	728	34	AB027581	AB027581 Parides n
c 44	18	3.4	813	34	MEU50216	U50216 Mytilus edu
c 45	18	3.4	1131	12	MMCALR	X73985 M.musculus

ALIGNMENTS

RESULT 1

HSDJ60101/c

LOCUS

DEFINITION

Human DNA sequence

sequence.

ACCESSION

33458 bp

DNA

sequence from clone 60101 on chromosome 20, complete

PRI

23-NOV-1999



```

/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
/map="ve016"
/clone="T20F21"
7..59
/rpt_family="(GA)n"
31..92
/note="exon predicted by xgrail, quality
marginal_shadowexon"
complement(87..145)
/rpt_family="POLY_A"
158..241
/rpt_family="(TAA)n"
complement(466..488)
/rpt_family="AT-rich"
514..565
/note="exon predicted by xgrail, quality
marginal_shadowexon"
complement(join<803..959,1112..1290,1352..>1360))
/gene="T20F21.1"
complement(<803..>1360)
/gene="T20F21.1"
/note="predicted by genefinder and genscan"
complement(join(803..959,1112..1290,1352..>1360))
/gene="T20F21.1"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAD15435.1"
/db_xref="GI:4263775"
/translation="MMATPFTAGVAVATAGVGIQAOAFKARPRPKIKFYE
GGFQPTMKREALLILGVRESVAEKVKAHRKVMVANHDPDAGGSHFLASKINEADV
MUGTKNSGSF"
1443..1580
/note="exon predicted by xgrail, quality
marginal_shadowexon"
1821..1964
/note="exon predicted by xgrail, quality
marginal_shadowexon"
complement(join(2016..2804,3330..>3402))
/gene="T20F21.2"
complement(2016..>3402)
/gene="T20F21.2"
complement(join(2161..2804,3330..>3402))
/gene="T20F21.2"
/note="unknown protein"
/codon_start=1
/protein_id="AAD15436.1"
/db_xref="GI:4263776"
/translation="WGRNALHLRSQSRLLSSFTSGYHRSSTAGSWSSVNPKVR
FVPSLSQSWASFQAKTRDDDEHKISGPQEKKEKDGWVYVYISSTIKVKLL
SLSTCLSVSLGPIFTFTSPGLNVMKGAVASTIFLSASTAALHWFVSPYVKLR
WQPGSDTFEVMNTWLTGTFSAKTLAFSDIRYPTQRPYVSFKADGNFYVDADHCPNK
ALLARLTPPKDAHDSAFKNL"
3528..3571
/note="exon predicted by xgrail, quality marginal"
complement(3614..3639)
/rpt_family="POLY_A"
join(<4033..4390,4472..4567,4643..4735,4808..4881,
4958..5224,5677..5804,5890..5985,6073..6178,6264..6471)
/gene="T20F21.3"
<4033..6471
/gene="T20F21.3"
join(4033..4390,4472..4567,4643..4735,4808..4881,
4958..5224,5677..5804,5890..5985,6073..6178,6264..6404)
/gene="T20F21.3"
/codon_start=1
/product="putative serine carboxypeptidase II"
/protein_id="AAD15437.1"
/db_xref="GI:4263777"
/translation="MARLLLFLLHLYASCSSRHEQKDRIFHLPCEPNDVGFSH
FSGYITVNESAGRALFYWLTPSPENPEKPLVLMNGPGCSSVAYGAEEIGPER
INPDGKTLVHNYSNKNLANLLFLESPAGVGFYSNTVSDLYTAGDQRTAEADYFLV
KWFERFPOYKHREYIAGESYAGHYVVPOLSOIYVEKRNPAINFKGFIQVNAVDDYHD
IVGLFEYWAHGLISDLIYHNLITCEGSESHPSKCTKAMEADLEQGNIDPYSIY
TVTCKKEAALRSRFSRVHRPMMRAYDPCTEKYSMTGFNSPEVQKAMHANTGLAYP
WKGSDIVGEKWDSPLSMLPIYKELIAAGRIWVFGSDTSDVPTITRYSIRALKL
OPLSKWYPMNDGQGVGSQVYKGLTLVTIHGAGHEVFLFRPRRAFLFQSLDNKPL
PM"
complement(4042..4094)
/rpt_family="(GAA)n"
complement(6713..6736)
/rpt_family="AT-rich"
join(<7685..8015,8104..8199,8286..8378,8910..8995,
9073..9545,9694..9970)
/gene="T20F21.4"
<7685..9970
/gene="T20F21.4"
join(7685..8015,8104..8199,8286..8378,8910..8995,
9073..9545,9694..9958)
/gene="T20F21.4"
/codon_start=1
/product="putative serine carboxypeptidase II"
/protein_id="AAD15438.1"
/db_xref="GI:4263778"
/translation="MVIALDVSSDDAKEOKMKDKIISLPQPPNLNFSQSGVTVV
DPAAGRALFYWLTEAPRSGTKPLVIMLNGGSCSIAYGASEEVPFRVNPDPGTILR
NLTAWKVAVNVLFDSPAGVGFSTINTSDELTGDRGTGATDARFLVWLERPEY
KERFYIAGESYAGHYIPELAQLIVNRNKGAKNPTINLKGILMGLNPLVDDYNDKGMR
DYWNHGLISDESYNDLTKWCLNSILFPLKNCNALNSALSEFGDIDPYNINSPACT
THASNEWMAQWRYGNDCEVGYTRKYMNPDNPVNHKSFHARLNGSTPWTGCSRVRKN
WKDSPKMLPIIKLLOAHLRIWIFSGSDAVLPLSGTRHSINAKMLKSSKRWYPWH
SHGLVGGSVYDEGLLTYTVRAAGHEVPLSQPLALFLTFHLANHSLPSPS"
complement(10090..10112)
/rpt_family="AT-rich"
complement(10160..10197)
/note="exon predicted by xgrail, quality
excellent_shadowexon"
11858..>12078)
/gene="T20F21.5"
complement(<10373..>12078)
/gene="T20F21.5"
/note="predicted by genscan"
complement(join(10373..10547,11083..11294,11641..11773,
11858..12078))
/gene="T20F21.5"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAD15439.1"
/db_xref="GI:4263779"
/translation="MSYLGVGSPGVSGSTTKMLIDRKVRYTELILRLCYVLALY
AAILIATDVREIFWIKKATDKMLVLLVYVNGIAAGSYLQVAVRCVVGMLKGR
VLFSPKLAWAIFFDQAVAYLCVAGAAQAQSAFALKEPELOWMKICNMVKGFCNQ
VGEIGTASALFACIGMLVISCISAFGVFRLLKRIEGRIVYNGQDHGEGGFGDGFVAY
VRVGFRCNTNRTEDRLSRNRLYQILR"
12202..12237
/note="exon predicted by xgrail, quality
excellent_shadowexon"
12634..12712
/note="exon predicted by xgrail, quality

Query Match 3.8%; Score 20; DB 8; Length 107436;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 387 tgattttgaaatctttgtaa 406
|||||
Db 68937 TGATTTGAAATCTTTGTAA 68956

RESULT 3
AC004932/c
LOCUS
DEFINITION Homo sapiens clone D0943F02, *** SEQUENCING IN PROGRESS ***, 3
unordered pieces.

```

AC004932 152143 bp DNA HTG 12-JUN-1998  
 LOCUS  
 DEFINITION Homo sapiens clone D0943F02, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 3  
 unordered pieces.

```

ACCESSION AC004932
VERSION AC004932.1 GI:3213065
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Waterston,R.H.
JOURNAL The sequence of Homo sapiens clone
REFERENCE 2 (bases 1 to 152143)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1504: contig of 1504 bp in length
* 1505 1522: gap of unknown length
* 1523 46957: contig of 45435 bp in length
* 46958 46975: gap of unknown length
* 46976 152143: contig of 105168 bp in length.
FEATURES
source
1..152143
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DJ0943F02"
BASE COUNT 41637 a 34439 c 33353 g 42678 t 36 others
ORIGIN

Query Match 3.8%; Score 20; DB 33; Length 152143;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 cttctgtcacaacaaaaa 526
|||||
Db 9837 CTTTCTGTCRAAAAAA 9818

RESULT 4
AC009560 173338 bp DNA HTG 29-SEP-1999
LOCUS Homo sapiens chromosome 15 clone 219_B_17 map 15, *** SEQUENCING IN
DEFINITION PROGRESS ***, 14 unordered pieces.
ACCESSION AC009560
VERSION AC009560.2 GI:5932623
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 15, clone 219_B_17
REFERENCE 2 (bases 1 to 173338)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., DeArillano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,

```

```

Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Torriella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 29, 1999 this sequence version replaced gi:5788084.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 9578: contig of 9578 bp in length
* 9579 12476: contig of 2898 bp in length
* 12477 16383: contig of 3907 bp in length
* 16384 19912: gap of unknown length
* 19913 25672: contig of 3529 bp in length
* 25673 32436: contig of unknown length
* 32437 40053: contig of 7617 bp in length
* 40054 51295: contig of 11242 bp in length
* 51296 63757: contig of 12462 bp in length
* 63758 76574: contig of unknown length
* 76575 91924: contig of 15350 bp in length
* 91925 117907: contig of 25983 bp in length
* 117908 143713: contig of 25806 bp in length
* 143714 173338: gap of unknown length
* 173339 29625: contig of 29625 bp in length.

```

TITLE  
JOURNAL  
COMMENT

FEATURES  
source

BASE COUNT 51440 a 35384 c 35491 g 49643 t 1380 others  
ORIGIN

Query Match 3.8%; Score 20; DB 41; Length 173338;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 aaagatgtgaactttgtc 231  
|||||  
Db 161601 AAAGATGTGAACITTTGTC 161620

RESULT 5  
AF061577  
LOCUS AF061577 1027 bp mRNA PLN 09-MAY-1998

REFERENCE 2 (bases 1 to 2165)

Sodergren, E., Hardham, J.M., McLeod, M.P., Salzberg, S., Peterson, J.,

REFERENCE 2 (bases 1 to 2165) .

Khalak, H., Richardson, D., Howell, J.K., Chidambaram, M.,  
 Otterback, T., McDonald, L., Artlich, P., Bowman, C., Cotton, M.D.,  
 Fujii, C., Garland, S., Hatch, B., Horst, K., Roberts, K., Watthey, L.,  
 Weidman, J., Smith, H.O. and Venter, J.C.  
 Direct Submission  
 Submitted (06-NAR-1998) The Institute for Genomic Research, 9712  
 Medical Center Dr. Rockville, MD 20850, USA  
 Location/Qualifiers  
 i. 10808  
 /organism="Treponema pallidum"  
 /db\_xref="taxon:160"  
 complement(85..363)  
 /gene="TP0597"  
 complement(85..363)  
 /gene="TP0597"  
 /note="hypothetical protein; identified by Glimmer;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="T. pallidum predicted coding region TP0597"  
 /protein\_id="AAC65380.1"  
 /db\_xref="GI:3322901"  
 /translation="MRRILSYKDEDIDAHMVMRRSLVNLGKALVGHVQSOQVYR  
 MGRYALPKVAITSGVSVNVYDGSVYEGDGLKATEFIRTEF"  
 complement(327..2144)  
 /gene="TP0598"  
 complement(327..2144)  
 /gene="TP0598"  
 /note="hypothetical protein; identified by Glimmer;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="T. pallidum predicted coding region TP0598"  
 /protein\_id="AAC65581.1"  
 /db\_xref="GI:3322902"  
 /translation="MGCCTGLFADSVKRAKDVTEPARFPYPSQKSTLESARSSTS  
 ESENASSVPSHQELAPDSAAPNVSLPAPPERKQGTAVHGAENVTRAGVSP  
 REVGLTKTLAASDHTFAAGNDGFLTOYTPDKPDWQITPVSIKCAVHPDARI  
 AVYEDGNRYHRVSVNNWTEKILFAKSTASVSLVIVGSLVGTASREGVTVL  
 DSGSTVSLFSEEPGVLLTAGPRLVLSYAESGRLLTVYDSKKTIVKRLLENLLS  
 PMLINGHALVQRQVIVIQSSGAVLTETPARSAFAHTSDSLVPWTEPAELKY  
 HWRKKAQRADFDPNARITISACSVTRVIVGTDRGILYELQGDGDRVTRALN  
 GQRVYASDVHGADGAYFLADGSLYHSMASGPPYRLVGVKGTREPLPYRDGRIWVS  
 AKGETYFLCAQKTSQRMVYRARSTVSGSVYGRMLVITPEFGSVVDIERGIVR  
 FHKAIGMODSLITDQVAVQSGLOPLVLLHMRGENYVQTRWEAICLGVAHDTQHV  
 YFSLDTNAGTTDLHFVCNSNPQKVLCDASSLIRMAI"  
 complement(2181..2837)  
 /gene="TP0599"  
 complement(2181..2837)  
 /gene="TP0599"  
 /note="hypothetical protein; identified by Glimmer;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="T. pallidum predicted coding region TP0599"  
 /protein\_id="AAC65582.1"  
 /db\_xref="GI:3322903"  
 /translation="MMQLRCACERVEDIEHETVLSLDEHPFEVARIQQGDFLSQCPA  
 CGARTRAKTEFVWAKNVHLLVPERLRCLAFACMGMSDGSADFCEPFLRE  
 HQPTVIGYELADVAAILADNLPVEAFAFFVLEGAPHGLGDKRVSFCFERCVGDTG  
 SRVMEHLVIREQQTALMPVPMNVYVERERKQAELEALYVGAYLSYKNVFTDA  
 complement(2844..4196)  
 /gene="TP0600"  
 complement(2844..4196)  
 /gene="TP0600"  
 /note="similar to GB:U42023 SP:P44936 PID:1006037  
 PID:1221023 PID:1205165 percent identity: 32.94;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="zinc protease, putative"  
 /protein\_id="AAC65573.1"

/db\_xref="GI:3322894"  
 /translation="MKILIIIGVVVLGIYVLFHELGHFVAALWCRVEVLSFVSGMGPVLT  
 FRKFKGTEYRLSMLPIGGICMGKEQAFQALDQKLSRIPEFGSLFVAGPLKRMGI  
 FAGPLANLVAMVVALVLSALGSRVHTFGNRISPVYVYDSDNSPARVGLQDGTI  
 LRIGDQPIRYFSDFIKIVYSHQARALPEVIERRGQLMHVITTPDRDAHTGMRGVGIYH  
 YVPLVAAVADHGAASRAGLEPEDKILAVAGRRVQHAVQVLLALLKEPRKSVLTVLR  
 SGKRRYHTIALVRENGAIDVGIEMKHTVVIPTGTSFASVRAGIETLRMCVLTVK  
 IGMFLFRGLQFOQAISGPLRITHVIGDVAGHGQFQSELTGLSQLCEFVALVYCSLFIMN  
 LLPILDGLLILFACVLELWQRSIHPRVLYLQFVGFALFLCAFWNDVNFLEH  
 complement(4193..5344)  
 /gene="TP0601"  
 complement(4193..5344)  
 /gene="TP0601"  
 /note="similar to PID:1001492 PID:1001556 percent  
 identity: 39.04; identified by sequence similarity;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAC65574.1"  
 /db\_xref="GI:3322895"  
 /translation="MFWDRCMSVRRVVVIGITSGISGAALKLLRRFPDRFLVWAGSG  
 HRQTEYARALARFSLSDITMTGSCSEQRARIKRLSSCEAEVYVNGTAGAGLGA  
 SLEVLKTRCTIALANKESVVLAAALLHAAARESGATVPVDSEHAAIFQIIAAHGAHA  
 VAQVLTASGSPFRTKECLAHVTVEDLQHPTRMGKISVDSATLANKALRVIA  
 VQFPRIPIVDRVTVVHPQSIHVHALVQCHSETVAQLSVPMASPLLYALLVPPAPPAY  
 QTPLDTSGLSLHEPPEPRVDPLLRMGFDVARAQRAYPIAFNANAEVRAFQQRNI  
 GFLDAHVTAAQLQEDWRAIPQTFEEVYACDTRAMCARTICIAQRWHER"  
 complement(5320..6183)  
 /gene="TP0602"  
 complement(5320..6183)  
 /gene="TP0602"  
 /note="similar to GB:AE000783 percent identity: 32.47;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="phosphatidate cytidylyltransferase (cdsa)"  
 /protein\_id="AAC65575.1"  
 /db\_xref="GI:3322896"  
 /translation="MSAEIKRLIFFGVPTILMYVAAHPAHFLAFLHVLVFSVWGA  
 VWEMHAMVSRMCTYPLVLLPLSLVLYCYAALWQPARGAESVLFIGALGTLMSV  
 FTVELYSFASFENALERMASALLVLYPGISLFFSLTRWRHAEIALVIFLMAV  
 TDCSQAFCGLTGMVNNRGLIPASPKSIAGFIQGGFAGSVGACGFCGSLVFGSVTSL  
 GMLMGVALVGLTAIVGDLVSEYMKRSQAVKDSGFFTPGREGIMDNLDSLAPSLGTFY  
 IACEFCGIAAV"  
 complement(6180..6866)  
 /gene="TP0603"  
 complement(6180..6866)  
 /gene="TP0603"  
 /note="similar to GB:AE000783 percent identity: 42.48;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAC65576.1"  
 /db\_xref="GI:3322897"  
 /translation="MQVAIIMDNGRWAEERGLRRSAGHRRGLQTAREIVAALCAIR  
 VPVTVYFSTENKRSHEVHFLMNLIRWLYLKEMSFYVHAIRVYVHLGCAQLPPD  
 VRSQIEVYVDSRHRGTVALAINVGGDEILRAVKVLCSTPCDGLLEAEARGA  
 CLDAPOLPSVDFLIRTGGQRMNSFLLWQSAEAFFTDILWPDFDFEDMLRALDEIR  
 LRTRTGGGLE"  
 complement(6872..7423)  
 /gene="TP0604"  
 complement(6872..7423)  
 /gene="TP0604"  
 /note="similar to GB:AL009126 percent identity: 46.29;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ribosome recycling factor"  
 /protein\_id="AAC65577.1"  
 /db\_xref="GI:3322898"



```
/translation="MGTAECYEQMKKSLSLALQEGFNTLRTERTATAHLLDQITVDYIQ
OPTALSVAVSVPEARLITIQPDWKILLADIERAILKSLSLNPSNDGKIVRIUVIPP
LTQRRRELVRQALAEQARVAIRNIRREGIEAKRGHKEGLLEDALKAAEAFQK
ATDASVADVARYLAEEKDILEG"
complement(7527. .8399)
/gene="TP0605"
CDS
complement(7527. .8399)
/gene="TP0605"
/translation="similar to SP:P34828 GB:X53651 PID:296031 percent
identity: 37.66; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="translation elongation factor TS (tsf)"
/protein_id="AAC65578.1"
/db_xref="GI:3322899"
/translation="MEIAARDVKSRLDKTGAGMECKRALQECAGDALCAEKYLKRG
LAATENRRGATAGVIVIKARHAEAGCAGSASAVMVELCETDFVAKNAEFIALER
IAQVLEHAYTEVQNVLDMVYDLATRVRENSLTRALLRAGSAGAQYLISHVHPD
KKTGVLSFSDADPDLRSVRAFAYDCCLHAAAYTPRVYRAEDVPAEYVREQREVF
QAHVASLQKPAHVRESIVQGLKELAEICFLKQPFVKDDKLSYKKAEGYARAGGA
LRFQALYQLGVQ"
complement(8490. .9365)
/gene="TP0606"
CDS
complement(8490. .9365)
/gene="TP0606"
/translation="similar to GB:AE000783 percent identity: 58.75;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
```

Query Match 3.6%; Score 19; DB 2; Length 10808;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 ttcttcattcttctgctcat 73  
|||||  
Db 7386 TCTTCATCTCTGCTCAT 7404

```
RESULT 8
AF094481 19839 bp DNA PRI 12-JAN-1999
LOCUS Homo sapiens trinucleotide repeat DNA binding protein p20-CGGBP
DEFINITION Homo sapiens trinucleotide repeat DNA binding protein p20-CGGBP
ACCESSION AF094481
VERSION AF094481.1 GI:4140681
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Mueller-Hartmann, H., Naumann, F., Deissler, H., Schmitz, B. and
Doerfler, W.
TITLE The 20 kDa 5'-(CGG)n-3'-binding protein is a nuclear protein
targeted to specific human chromosomes and can affect the activity
of the FMR1-promoter
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 19839)
AUTHORS Naumann, F.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-1998) Department for Medical Genetics and
Virology, Institute of Genetics, University of Cologne, Weyertal
121, Cologne 50931, Germany
FEATURES
source
1.19839
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p11.1-p12"
/cell_line="HeLa"
repeat_region 1. .67
```

```
/rpt_family="alu"
/rpt_type="dispersed
326. .607
/rpt_family="alu"
/rpt_type="dispersed
join(7396. .7544,8136. .8340,9067. .9168,10360. .10946)
/gene="CGGBP"
/product="trinucleotide repeat DNA binding protein
p20-CGGBP"
7396. .10946
/gene="CGGBP"
10383. .10886
/gene="CGGBP"
/function="binds sequence-specific to the double-stranded
triplet repeat 5-d(CGG)n-3"
/notes="binding is severely inhibited by cytosine-specific
DNA-methylation"
/codon_start=1
/product="trinucleotide repeat DNA binding protein
p20-CGGBP"
/protein_id="AAD04161.1"
/db_xref="GI:4140682"
/translation="MERFVTPADPARNRSKTALVTPLDRTVEFGELHEDGGKLFCT
SCNVLNHRKSAISDLKSKTHKRAEFEQNVKQRPRTASLQCNSTAQTEKVS
VIQDFVAKLEANIPLEKADHPAVRAFLSRHVKNGSIPKSDQLRRAYLPDGYENQ
LLNSQDC"
/misc_feature 10587. .10634
/gene="CGGBP"
/notes="encodes nuclear localization signal"
14878. .15163
/rpt_family="alu"
/rpt_type="dispersed
15701. .15980
/rpt_family="alu"
/rpt_type="dispersed
16496. .16776
/rpt_family="alu"
/rpt_type="dispersed
16969. .17175
/rpt_family="alu"
/rpt_type="dispersed
17821. .18086
/rpt_family="alu"
/rpt_type="dispersed
19151. .19491
/notes="similar to transposon-like element of GenBank
Accession Number U19976"
19793. .19839
/rpt_family="alu"
/rpt_type="dispersed
6102 a 3874 c 3683 g 6180 t
BASE COUNT
ORIGIN
Query Match 3.6%; Score 19; DB 40; Length 19839;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 205 acctccaaaagatgtgaaa 223
|||||
Db 1820 ACCTCCAAAAGATGTGAAA 1838
RESULT 9
HUMCC1S149/c
LOCUS HUMCC1S149 34379 bp DNA PRI 10-MAR-1995
DEFINITION Homo sapiens cosmid clone HDAB (1S149) insert DNA, complete cosmid.
ACCESSION M63005
VERSION M63005.1 GI:179981
KEYWORDS Human DNA, cosmid clone HDAB (1S149), pCOS2 vector library of J.F.
Gusella and M.E.MacDonald.
SOURCE Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

# REFERENCE AUTHORS

McCombie, W.R., Martin-Gallardo, A., Gocayne, J.D., FitzGerald, M.G.,  
Dubnick, M., Kelly, J.M., Castilla, L., Liu, L.-I., Wallace, S.,  
Trapp, S., Tagle, D.A., Whaley, W.L., Cheng, S., Guse, J.,  
Frischauf, A.-M., Poustka, A., Lehrach, H., Collins, F.S.,  
Kerlavage, A.R., Fields, C. and Venter, J.C.

Expressed genes, Alu repeats and polymorphisms in cosmids sequenced  
from chromosome 4p16.3

Nature Genet. 1 (5), 348-353 (1992)

93251024

# FEATURES

Location/Qualifiers

1..34379

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="1S149"

7822 a 8904 c 9450 g 8203 t

Chromosome 4p16.3.

# Query Match

Best Local Similarity 3.6%; Score 19; DB 9; Length 34379;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

# Qy

361 aaataaagacacaagaaaa 379

# Db

3233 AAATAAGACACAGAAAA 3215

# RESULT 10

AC006646

LOCUS

AC006646 38702 bp DNA HTG 23-FEB-1999

Caenorhabditis elegans clone F58D5, \*\*\* SEQUENCING IN PROGRESS \*\*\*,

1 unordered pieces.

AC006646

VERSION AC006646.1 GI:4263292

HTG; HTGS\_PHASE1.

SOURCE

Caenorhabditis elegans.

ORGANISM

Eukaryota; Metazoa; Nematoda; Secernentes; Rhabditia; Rhabditida;

Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 38702)

Waterston, R.H.

The sequence of Caenorhabditis elegans clone

Unpublished

2 (bases 1 to 38702)

Waterston, R.H.

Direct Submission

Submitted (23-FEB-1999) Genome Sequencing Center, Washington

University School of Medicine, 444 Forest Park Parkway, St. Louis,

MO 63108, USA

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 1 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 38702: contig of 38702 bp in length.

Location/Qualifiers

1..38702

/organism="Caenorhabditis elegans"

/db\_xref="taxon:6239"

/clone="F58D5"

12205 a 7031 c 6828 g 12638 t

Chromosome 4p16.3.

# Query Match

Best Local Similarity 3.6%; Score 19; DB 33; Length 38702;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

# Qy

508 ttctgtcaaaaaa 526

# Db

37026 TTTCTGTCAAAAA 37044

# RESULT 11

CEH04109/c

LOCUS

CEH04109/c

DEFINITION

Caenorhabditis elegans chromosome II clone H04109, \*\*\* SEQUENCING

IN PROGRESS \*\*\*; in unordered pieces.

ACCESSION

Z92848

VERSION

Z92848.1 GI:1888365

HTG; HTGS\_PHASE1.

SOURCE

Caenorhabditis elegans.

ORGANISM

Eukaryota; Metazoa; Nematoda; Secernentes; Rhabditia; Rhabditida;

Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 40873)

Sulston, J.

Direct Submission

Submitted (11-MAR-1997) Nematoe Sequencing Project, Sanger Centre,

Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,

Washington University, St. Louis, MO 63110, USA. E-mail:

jes@sanger.ac.uk or rwenematode.wustl.edu

IMPORTANT: This sequence is unfinished and does not necessarily

represent the correct sequence. Work on the sequence is in progress

and the release of this data is based on the understanding that the

sequence may change as work continues. The sequence may be

contaminated with foreign sequence from E.coli, yeast, vector,

phage etc. Order of segments is not known; 800 n's separate

segments.

\* NOTE: This is a 'working draft' sequence.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

Location/Qualifiers

1..40873

/organism="Caenorhabditis elegans"

/db\_xref="taxon:6239"

/chromosome="II"

/clone="H04109"

9720 a 4810 c 4691 g 9562 t 12090 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 3.6%; Score 19; DB 32; Length 40873;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

506 ttctgtcaaaaaa 524

Db

26654 TCTTCTGTCAAAAA 26636

# RESULT 12

CH19F15314/c

LOCUS

CH19F15314/c

DEFINITION

Homo sapiens DNA from chromosome 19p13.1 cosmid f15314, genomic

sequence.

ACCESSION

AD000091

VERSION

AD000091.1 GI:1905896

KEYWORDS

D19S11; chromosome 19; cytochrome P450; pseudogene.

SOURCE

Homo sapiens (clone: F15314) (clone\_lib: L L19NCO2 F2 chromosome

19-specific cosmid library) ss-DNA.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 41369)

Lamerdin, J.E., Stilwagen, S.A., Ramirez, M.H., McCready, P.,

Hoffman, S., Mohrenweiser, H. and Carrano, A.V.

Identification of a CYP4F gene family at the D19s11 locus on

19p13.1

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 41369)
AUTHORS Lamerdin,J.E.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-1996) J.E. Lamerdin, Human Genome Center,
Lawrence Livermore National Laboratory, 7000 East Ave, Livermore,
CA, USA, 94551 jane@acct.llnl.gov ow@tornak.llnl.gov
COMMENT
GSD: S:1079875.
constructed at LLNL from flow-sorted chromosomes
from hybrid UV5HL9-5B, which carries chromosome 19 as its only
human chromosome.
FEATURES
source Location/Qualifiers
1. 41369 Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/cell_line="UV5HL9-5B"
/cell_type="fibroblast"
/clone="F15314"
/clone_lib="L L19NCO2 F2 chromosome 19-specific cosmid
library"
repeat_region complement(783..1063)
/feature="repeat match = HSAL06523; putative"
/rpt_family="Alu"
repeat_region complement(1029..1072)
/feature="repeat match = HSAL05178; putative"
/rpt_family="Alu"
misc_feature 1396..1464
/feature="similarity: sp|Q08477; similarity to CPF3 (human);
putative"
misc_feature 1401..1461
/feature="predicted exon, grill2exons_human_1.3; frame=2,
forward strand, quality=marginal; putative"
misc_feature 1405..1464
/feature="similarity: sp|P33274; similarity to CP4F (rat);
putative"
exon 1408..1461
/feature="Human Leukotriene B4 omega-hydroxylase family
pseudogene, exon 4; putative; does not fit consensus"
/number=4
misc_feature 1544..1681
/feature="similarity: sp|Q08477; similarity to CPF3 (human);
putative"
misc_feature 1544..1681
/feature="similarity: sp|P33274; similarity to CP4F (rat);
putative"
misc_feature 1553..1680
/feature="predicted exon, grill2exons_human_1.3; frame=0,
forward strand, quality=excellent; putative"
exon 1553..1680
/feature="Human Leukotriene B4 omega-hydroxylase family
pseudogene, exon 5; putative; does not fit consensus"
/number=5
repeat_region /pseudo
1985..2244
/feature="repeat match = HSAL00518; putative"
/rpt_family="Alu"
repeat_region 2136..2275
/feature="repeat match = HSAL00299; putative"
/rpt_family="Alu"
misc_feature 2395..2416
/feature="predicted exon, grill2exons_human_1.3; frame=0,
forward strand, quality=good; putative"
misc_feature 2504..2586
/feature="predicted exon, grill2exons_human_1.3; frame=2,
forward strand, quality=good; putative"
repeat_region complement(2910..2977)
/feature="repeat match = HSAL04183; putative"
/rpt_family="Alu"
repeat_region complement(2924..3208)
/feature="repeat match = HSAL06258; putative"
/rpt_family="Alu"
exon 3527..3648
/feature="Human Leukotriene B4 omega-hydroxylase family
pseudogene, exon 6; putative; does not fit consensus"
/number=6
misc_feature 3527..3648
/feature="predicted exon, grill2exons_human_1.3; frame=1,
forward strand, quality=excellent; putative"
misc_feature 3528..3650
/feature="similarity: sp|Q08477; similarity to CPF3 (human);
putative"
misc_feature 3528..3650
/feature="similarity: sp|P33274; similarity to CP4F (rat);
putative"
exon 4378..4648
/feature="Human Leukotriene B4 omega-hydroxylase family
pseudogene, exon 7; putative; does not fit consensus"
/number=7
misc_feature 4380..4649
/feature="similarity: sp|P33274; similarity to CP4F (rat);
putative"
misc_feature 4380..4649
/feature="similarity: sp|Q08477; similarity to CPF3 (human);
putative"
misc_feature 4387..4648
/feature="predicted exon, grill2exons_human_1.3; frame=1,
forward strand, quality=excellent; putative"
misc_feature 4684..4778
/feature="predicted exon, grill2exons_human_1.3; frame=0,
forward strand, quality=marginal; putative"
misc_feature 4986..5052
/feature="predicted exon, grill2exons_human_1.3; frame=2,
forward strand, quality=excellent; putative"
exon 4986..5052
/feature="Human Leukotriene B4 omega-hydroxylase family
pseudogene, exon 8; putative; does not fit consensus"
/number=8
misc_feature 4987..5055
/feature="similarity: sp|P33274; similarity to CP4F (rat);
putative"
misc_feature 4987..5055
/feature="similarity: sp|Q08477; similarity to CPF3 (human);
putative"
misc_feature 5244..5373
/feature="predicted exon, grill2exons_human_1.3; frame=1,
forward strand, quality=excellent; putative"
exon 5244..5373
/feature="Human Leukotriene B4 omega-hydroxylase family
pseudogene, exon 9; putative; does not fit consensus"
/number=9
misc_feature 5244..5375
/feature="similarity: sp|P33274; similarity to CP4F (rat);
putative"
misc_feature 5244..5375
/feature="similarity: sp|Q08477; similarity to CPF3 (human);
putative"
repeat_region complement(5596..5707)
/feature="repeat match = HSAL05171; putative"
/rpt_family="Alu"
repeat_region complement(5625..5924)
/feature="repeat match = HSAL02250; putative"
/rpt_family="Alu"
misc_feature 6290..7080
/feature="ss region (gap in top strand); putative"
repeat_region complement(6665..6950)
/feature="repeat match = HSAL00697; putative"
/rpt_family="Alu"
repeat_region complement(6673..6879)
/feature="repeat match = HSAL05100; putative"
/rpt_family="Alu"

```

```
repeat_region complement(6884..6967)
/note="repeat match = HSAL06328; putative"
/rpt_family="Alu"
repeat_region complement(7298..7445)
/note="repeat match = HSAL04736; putative"
/rpt_family="Alu"
repeat_region complement(7331..7620)
/note="repeat match = AL0; putative"
/rpt_family="Alu"
repeat_region complement(7548..7639)
/note="repeat match = HSAL02061; putative"
/rpt_family="Alu"
repeat_region complement(7739..8057)
/note="repeat match = HSAL06256; putative"
/rpt_family="Alu"
repeat_region complement(7837..8068)
/note="repeat match = HSAL00736; putative"
/rpt_family="Alu"
misc_feature 9050..9100
/note="ss region (gap in bottom strand); putative"
repeat_region complement(9110..9354)
/note="repeat match = HSAL04630; putative"
/rpt_family="Alu"
repeat_region complement(9173..9712)
/note="repeat match = HSAL01827; putative"
/rpt_family="Alu"
repeat_region complement(9677..9723)
/note="repeat match = HSAL03456; putative"
/rpt_family="Alu"
repeat_region complement(9818..9930)
/note="repeat match = HSAL00794; putative"
/rpt_family="Alu"
repeat_region complement(9841..9977)
/note="repeat match = HSAL01882; putative"
/rpt_family="Alu"
repeat_region complement(9873..9989)
/note="repeat match = HSAL05507; putative"
/rpt_family="Alu"

Query Match 3.6%; Score 19; DB 10; Length 41369;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 atatgttgagacccaag 136
|||||
Db 13787 ATATGTGAGACCTCAAG 13769

RESULT 13
CEY54G9A/c
LOCUS CEY54G9A 42141 bp DNA INV 02-SEP-1999
DEFINITION Caenorhabditis elegans cosmid Y54G9A, complete sequence.
ACCESSION AL032648
VERSION AL032648.1 GI:3810715
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 42141)
AUTHORS Wilson,R., Alnough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
O'Callaghan,M., Parsons,R., Percy,C., Rifkin,L., Roopra,A.,
Saunders,D., Showkneen,R., Smalton,N., Smith,A., Sonnenhammer,B.,
Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
Wilkinson-Sproat,J., and Wohlman,P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
```

```
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 42141)
AUTHORS Smye,K.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-1998) Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwnematode.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using
predictions from Genefinder (P. Green, U. Washington), and other
available information.
for a graphical representation of this sequence and its analysis
see:-
http://webc.sanger.ac.uk/cgi-
bin/display?db=wormacsclass=Sequence object=y54g9a
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone Y54G9A.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone R0689 is at 42042 in this sequence. The
true right end of clone Y48E1A is at 100 in this sequence. The
start of this sequence (1..100) overlaps with the end of sequence
Z93392.
The end of this sequence (42042..42141) overlaps with the start of
sequence Z83237.
FEATURES
Location/Qualifiers
1. 42141
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="II"
/clone="Y54G9A"
5370..6996
/gene="Y54G9A.1"
join(5370..5395,5447..5542,5638..5693,6820..6855,
6911..6996)
/gene="Y54G9A.1"
/note="predicted using Genefinder"
/codon_start=1
/protein_id="CAA21697.1"
/db_xref="GI:3881112"
/translation="MTRKRSIRRHISIFMRLKMTSTYFFGGSFVSGVATYANNWVF
SLKFGFTNSFVLETFVYIEIQVTKRHKSEAHSTFTLRNPISFFTLSHRK"
complement(11862..13652)
/gene="Y54G9A.2"
complement(join(11862..12065,13429..13499,13553..13652))
/gene="Y54G9A.2"
/note="predicted using Genefinder"
/codon_start=1
/protein_id="CAA21696.1"
/db_xref="GI:3881111"
/translation="MPTCLRAYLQRIISFSLRRVQNTIAYLPSADEATGTELKWH
GAACSPSLKQDQSKLEYPQIVDSRPEKECENKENVVNNEMIVSVREKESAA
REGMDTPTSAHELKSNCSKEH"
14551..18795
/gene="Y54G9A.3"
join(14551..14718,15074..15204,15248..15390,15536..16325,
17316..17705,17789..17891,18583..18795)
/gene="Y54G9A.3"
/note="similar to Ion transport proteins"
/codon_start=1
/protein_id="CAA21699.1"
/db_xref="GI:3881114"
/translation="MTYRAMYINCLERTPTGWKCFYHFSVFLVILICILSVLSTVEE
HSFAAEALLYLEIVLVFVSVEFVILWSAGCSKYIGVYGLKFKVTKPTLLGRFG
GRHLQLTRDLRHGSPFCGRKRDAGDIVSDSDAPRRSAGHWVEIRISKPIKFIKF
```

KSLKIGIFKFNANNFQFKLEFSKILKNKAGRVHWGSAFSSGDDPLDVCCEPK  
 LHQASFRLLGSVVFTHROELITLITLIGLGLIFSVFYVLAEDHIGVDGQAFTS  
 QADALWGVITMTTIGVDVVPQWGLRIVASCFSIFAISFPALPAGILGSGFALKVQ  
 QOKROKHFNQVITIGVQWGLRIVASCFSIFAISFPALPAGILGSGFALKVQ  
 ASSMSNNLTFRKLFKQSSLVNTRFKGSPSDVEMGELQOERLLRHNSDDE  
 KRIYRGADIDVIDETEPTPTPQGHSHVCLTDAHNAIARIRVKYFVARRE  
 QOAKRVDIVDIOYQOGLHNMVRIKELQRLDDQTLGPGQDQKGRSGKHGVPVIG  
 SRGLRLEQMSLSRQSVSSNLTALYRLMADRNSLTISPPALISRPVSPACLS  
 PRDLSTPSSISQSRSGSPGQ"  
 complement(19069..21893)  
 /gene="Y54G9A.4"  
 complement(join(19069..19173,19824..20201,20541..20805,  
 21116..21306,21813..21893))  
 /gene="Y54G9A.4"  
 /note="predicted using GeneFinder"  
 /codon\_start=1  
 /protein\_id="CAA21695.1"  
 /db\_xref="GI:3881110"  
 /translation="MSLELLKWLGVNMTIIFGLLPKIVISYLNTRSAIHQHS  
 LILFSFAGVGLSVCFDMLPDLCEAWESVQDNTYSDYFVQLIALHGFYVY  
 LTELSSVICHVGHSHNDPIMESNVTFPRARLATVGSIFNVGALPCKRSLN  
 YDDGEGPVRSIIFTSAFILHVFECFAFIQEDAVSVTIFGLIAHKAIVMSFLG  
 MKLARTRPRRSWIVILLILVALENVIGTAGILISSNNQTPKDTITAVLMSFSLG  
 TELYISFEILLAPERANNHNSILQWIASFGGFALLAVNMIAT"  
 29014..30387  
 /gene="Y54G9A.5"  
 join(29014..29234,29284..29413,29469..29601,30649..30739,  
 30789..30987)  
 /gene="Y54G9A.5"  
 /note="predicted using GeneFinder: cDNA EST EMBL:D71366  
 comes from this gene; cDNA EST EMBL:D74042 comes from this  
 gene; cDNA EST EMBL:D71691 comes from this gene; cDNA EST  
 EMBL:D74364 comes from this gene; cDNA EST EMBL:D66668  
 comes from this gene; cDNA EST EMBL:D70534 comes from this  
 gene; cDNA EST EMBL:C10326 comes from this gene; cDNA EST  
 EMBL:C11939 comes from this gene; cDNA EST EMBL:C07915  
 comes from this gene; cDNA EST EMBL:C09472 comes from this  
 gene; cDNA EST YK284d1.3 comes from this gene; cDNA EST  
 YK284d1.5 comes from this gene; cDNA EST YK641a11.3 comes  
 from this gene"  
 /codon\_start=1  
 /protein\_id="CAA21700.1"  
 /db\_xref="GI:3881115"  
 /translation="MRSRQARAYKIFAFSQHFLPRINALKAFIIDSAMKLSIFW  
 ISILLVVCDAETPDRLRDEFRALSRDYLMLKQSPNDLSLMAFYVNGSVAA  
 ASSYASSIDRKNQGTILEETCGSAETHGWCQPKGRIGILTARSENRTALAHFLCD  
 LDRPKSNHYRETKILNCAFLGSLIEMIALNWKIKTRAVPVHPDILNLTGKSQSK  
 YESDSFESITSSKRSKISNLRSTFCAGPKKFRITA"  
 32132..33514  
 /gene="Y54G9A.6"  
 join(32132..32365,33689..34326,35355..35514)  
 /gene="Y54G9A.6"  
 /note="predicted using GeneFinder: similar to WD domain,  
 G-beta repeat (2 domains); cDNA EST EMBL:D65403 comes from  
 this gene; cDNA EST EMBL:D68813 comes from this gene; cDNA  
 EST EMBL:D66576 comes from this gene; cDNA EST EMBL:D70404  
 comes from this gene; cDNA EST EMBL:C07632 comes from this  
 gene; cDNA EST EMBL:C09060 comes from this gene; cDNA EST  
 YK216f6.3 comes from this gene; cDNA EST YK216f6.5 comes  
 from this gene; cDNA EST YK219e3.5 comes from this gene;  
 cDNA EST YK222h2.3 comes from this gene; cDNA EST  
 YK222h2.5 comes from this gene; cDNA EST YK262b10.3 comes  
 from this gene; cDNA EST YK262b10.5 comes from this gene;  
 cDNA EST YK276b11.5 comes from this gene; cDNA EST  
 YK390h8.5 comes from this gene; cDNA EST YK399a8.5 comes  
 from this gene; cDNA EST YK416d7.5 comes from this gene;  
 cDNA EST YK459a2.5 comes from this gene; cDNA EST  
 YK483b4.5 comes from this gene; cDNA EST YK489g12.5 comes  
 from this gene; cDNA EST YK500a5.5 comes from this gene;  
 cDNA EST YK539b6.3 comes from this gene; cDNA EST  
 YK579a7.3 comes from this gene; cDNA EST YK610d7.3 comes  
 from this gene; cDNA EST YK611d9.3 comes from this gene;  
 cDNA EST YK613g8.3 comes from this gene; cDNA EST

YK652e3.3 comes from this gene; cDNA EST YK663c12.3 comes  
 from this gene"  
 /codon\_start=1  
 /protein\_id="CAA21698.1"  
 /db\_xref="GI:3881113"  
 /translation="MSTYQATVAAPNEFRVPPFVQISKVQFQREAGSRLAASG  
 WDGTCRYVEVGKLDIDSEKLVFTFKLLCTFAGYKNKAFGGVDHNVKLADIETNG  
 TQGLSHALAYRCMEFNPMSSLVSGGWSVKLMDARSYNGALIESVNVSSSVTAMDV  
 LKHTILVTGDKRIFMYDSRLKREPLQVRDPSLKYQTRAVQVFFTGAFVVSIEGRV  
 AVEYVDQSGEEMAKRKVAFKCHREKDTGTELHPVHTVAFHPKRYGTAFGAGDIYN  
 IWDPNFRKRIQLHKKFTSISLSEFNGDSQALATATSYQYEKEIDPSPLPNNSITIRH  
 ITDPSRPK"  
 join(complement(700..831),complement(471..641),  
 complement(101..406),complement(293392.1:15711..17576),  
 complement(293392.1:12939..14133),  
 complement(293392.1:11449..11770),  
 complement(293392.1:9008..9640),  
 complement(293392.1:5923..6151),  
 complement(293392.1:477..669),  
 complement(283219.1:35329..35495))  
 /gene="Y48E1A.1"  
 join(complement(700..831),complement(471..641),  
 complement(101..406),complement(293392.1:15711..17576),  
 complement(293392.1:12939..14133),  
 complement(293392.1:11449..11770),  
 complement(293392.1:9008..9640),  
 complement(293392.1:5923..6151),  
 complement(293392.1:477..669),  
 complement(283219.1:35329..35495))  
 /gene="Y48E1A.1"  
 /note="similar to DNA-directed RNA polymerase I largest  
 Query Match 3.6%; Score 19; DB 34; Length 42141;  
 Best Local Similarity 100.0%; Pred.No.13;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 506 tcttctgtctcaaaaaa 524  
 |||||  
 DB 6957 TCTTCTCTCAAAAAA 6939  
 RESULT 14  
 HSJ421I20 52597 bp DNA PRI 23-NOV-1999  
 LOCUS Human DNA sequence from clone 421I20 on chromosome Xq22.1-23,  
 DEFINITION complete sequence.  
 ACCESSION AL117327  
 VERSION AL117327.5 GI:6066122  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 52597)  
 AUTHORS Bird,C.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-OCT-1999) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerquest@sanger.ac.uk  
 COMMENT On Oct 19, 1999 this sequence version replaced gi:6002341.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information  
 on the WormPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C/elegans/wormpep/IMPORTANT: This  
 sequence is not the entire insert of clone 421I20. It may be  
 shorter because we only sequence overlapping sections once, or

Query Match 3.6%; Score 19; DB 9; Length 58864;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 18:50:59 ; Search time 250.34 Seconds  
(without alignments)  
525.689 Million cell updates/sec

Title: US-09-215-435-76

Perfect score: 526  
Sequence: 1 ctgcctgtctgtgtgcac.....ctttctgcaaaaaaaaaa 526

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 311585 seqs, 125096042 residues

Word size : 0

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	237	45.1	271	1 X51569	Human secreted pro
2	19	3.6	2074	1 X20666	Polynucleotide seq
3	18	3.4	249	1 Q60466	Human brain Exres
4	18	3.4	1381	1 R84482	Human secreted pro
5	18	3.4	1713	1 Q03827	Pyruvate oxidase (
6	18	3.4	1713	1 Q08597	Pyruvate oxidase (
7	17	3.2	245	1 X07195	Human endogenous r
8	17	3.2	245	1 X07520	Human endogenous r
9	17	3.2	245	1 X07520	Human secreted pro
10	17	3.2	299	1 X69328	mak TTC8 variable
11	17	3.2	312	1 X23242	Nucleic acid encod
12	17	3.2	318	1 X82857	EST clone AM235. N
13	17	3.2	330	1 X86315	3'-flanking sequen
14	17	3.2	472	1 T12399	Anti-CD19 antibody
15	17	3.2	723	1 T47733	Human CD30 binding
16	17	3.2	744	1 V11399	DNA encoding a hum
17	17	3.2	921	1 X51728	Breast cancer asso
18	17	3.2	985	1 X39624	Human secreted pro
19	17	3.2	1037	1 X04319	Human secreted pro
20	17	3.2	1052	1 X04397	Human secreted pro
21	17	3.2	1445	1 V42294	Human secreted pro
22	17	3.2	1505	1 T72177	Alzheimer's diseas
23	17	3.2	2188	1 Q79726	Mouse L5/3 tumour
24	17	3.2	2188	1 T62441	Mouse growth facto
25	17	3.2	2193	1 V71267	Syndecan interacti
26	17	3.2	2290	1 N90615	CDw32a cDNA. Rapid
27	17	3.2	2290	1 Q21169	Human CDw32a anti
28	17	3.2	2290	1 Q21174	Human CDw32a anti
29	17	3.2	2290	1 T14714	Human CDw32a anti
30	17	3.2	2290	1 T14709	Human CDw32a anti
31	17	3.2	2290	1 V81205	Human CDw52 anti
32	17	3.2	2290	1 V81210	Human CDw32a anti
33	17	3.2	2290	1 V63452	Human CDw32a anti
				1 V63447	Human CDw52 anti

34	17	3.2	3523	1 T93091	Human transcriptio
35	17	3.2	8160	1 V99559	Arabidopsis lysine
36	17	3.2	8174	1 Q13332	GDP-Fuc:beta-D-gal
37	17	3.2	8174	1 Q56908	DNA encoding a gly
38	17	3.2	8174	1 T61677	Human alpha(1,2)-f
39	17	3.2	50000	1 X23517	Human kidney amino
40	17	3.2	50000	1 X23517	Human kidney amino
41	17	3.2	53585	1 X20251	Borrelia burgdorfe
42	17	3.2	237326	1 V57903	Hereditary haemoch
43	16	3.0	210	1 T22245	Human gene signatu
44	16	3.0	421	1 X21198	Polynucleotide seq
45	16	3.0	531	1 X21189	Polynucleotide seq

## ALIGNMENTS

RESULT 1	
X51569	
ID X51569 standard; cDNA; 271 BP.	
AC X51569;	
DE 21-JUN-1999 (first entry)	EST SEQ ID NO:148.
DE Human secreted protein 5;	EST expressed sequence tag; diagnosis;
KW Human; secreted protein; EST;	expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping;	signal peptide;
KW upstream regulatory sequence; cytokine activity;	cell proliferation;
KW differentiation; haematopoiesis regulation;	tissue growth regulation;
KW reproductiv hormone regulation;	chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory;	tumour inhibition; ds.
OS Homo sapiens.	
PN WO9906549-A2.	
PD 11-FEB-1999.	
PF 31-JUL-1998; IB1231.	
PR 01-AUG-1997; US-905279.	
PA (GEST ) GENSET.	
PI Duclert A, Dumas Milne Edwards J, Lacroix B;	
DR WFI; 99-153779/13.	
DR P-PSDB; Y12791.	
PT New nucleic acids encoding human secreted proteins - obtained from	
PT cDNA libraries derived from testis, ovary, uterus and spleen tissue	
PS Claim 1: Page 253-254; 522pp; English.	
CC X51459 to X51691 represent 5' expressed sequence tags (ESTs) for human	
CC secreted proteins, and encode the proteins given in Y12681 to Y12913,	
CC respectively. The proteins given represent the signal peptide and an	
CC N-terminal fragment of a secreted protein. The nucleic acid sequences	
CC can be used for producing secreted human gene products. They can also	
CC be used to develop products for diagnosis and therapy. The proteins	
CC obtained may have cytokine activity, cell proliferation/differentiation	
CC activity, haematopoiesis regulating activity, tissue growth regulating	
CC activity, reproductiv hormone regulating activity, chemotactic/	
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/	
CC ligand activity, anti-inflammatory activity, tumour inhibition activity	
CC or other activities. The products can be used in forensic, gene therapy	
CC and chromosome mapping procedures. The sequences can also be used for	
CC obtaining corresponding promoter sequences. The nucleic acids encoding	
CC the signal peptide can be used for directing extracellular secretion of	
CC a polypeptide or the insertion of a polypeptide into a membrane, or	
CC importing a polypeptide into a cell.	
SQ Sequence 271 BP; 75 A; 57 C; 84 T;	

Query Match 45.1%; Score 237; DB 1; Length 271;  
Best Local Similarity 100.0%; Pred. No. 9.3e-106;  
Matches 237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1	ctgcctgtctgtgcacacgaagctgcccaagctgggattcttcttaagattcttc	60
Db		
QY 11	ctgcctgtctgtgcacacgaagctgcccaagctgggattcttcttaagattcttc	70
Db		
QY 61	attcttgcctcattgaataccctgttatttggtgggttaataaaatgcggagaagata	120
Db		
QY 71	attcttgcctcattgaataccctgttatttggtgggttaataaaatgcggagaagata	130
Db		

QY 121 tgtggagacctcaaatgaccccaattggacatgaatttggagctgctataagtt 180  
DB 131 TGTGGAGACCTCAGAGATCCCTGCAATTTGGACATGAATTTTGGAGCTGCTATGAAGTT 190  
QY 181 cacttttagatatttctacacagaaacctccaaaagtgtgaaacttttctctctcc 237  
DB 191 CACTTTAGATATTCTTACACAGAACTCCAAAGATGTGAACCTTTGCTCTCTCC 247

## RESULT 2

X20666/c  
ID X20666 standard; DNA; 2074 BP.  
AC X20666;  
DT 05-MAY-1999 (first entry)  
DE Polynucleotide sequence from the genome of *Treponema pallidum*.  
KW *Treponema pallidum* infection; syphilis; *Borrelia* infection; animal;  
OS enzyme production; ds.  
PN *Treponema pallidum*.  
PD W09859034-A2.  
PD 30-DEC-1998.  
PF 23-JUN-1998; U13041.  
PR 24-JUN-1997; US-050667.  
PA (HOMA-) HUMAN GENOME SCI INC.  
PI Fraser CM;  
DR WPI; 99-081273/07.  
PT New isolated *Treponema pallidum* nucleic acids - used to develop  
PT products for the detection, diagnosis, characterisation, prevention  
PT and therapy of *T. pallidum* infections, particularly syphilis  
PS Claim 1; Page 829-830; 1150pp; English.  
CC X20500-21243 represent polynucleotide sequences from the genome of  
CC *Treponema pallidum*. The sequences can be used for detection,  
CC diagnosis, characterisation, prevention and therapy for *T. pallidum*  
CC infections, particularly syphilis. They can also be used for detecting  
CC diseases related to *Borrelia* infections in animals, and for the  
CC production of biosynthetic products such as enzymes.  
SQ Sequence 2074 BP; 466 A; 456 C; 665 G; 480 T;

## Query Match

Best Local Similarity 3.68; Score 19; DB 1; Length 2074;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 ttcttcattcttctgtctcat 73  
DB 1342 TTCTTCATCTTCTGCTCAT 1324

## RESULT 3

Q60466/c  
ID Q60466 standard; cDNA; 249 BP.  
AC Q60466;  
DT 16-MAR-1994 (first entry)  
DE Human brain Expressed Sequence Tag EST02470.  
KW Gene transcription product; genetic markers; tagging; in vivo;  
KW transcription; mapping; locations; chromosomes; chromosomal; ss.  
OS Homo sapiens.  
PN W09316178-A.  
PD 19-AUG-1993.  
PF 12-FEB-1993; U01294.  
PR 12-FEB-1992; US-837195.  
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.  
PI Adams MD, Moreno RF, Venter CJ;  
DR WPI; 93-272882/34.  
PT Enriched oligonucleotides and corresp. sequences - used as  
PT markers for human genes transcribed in-vivo, facilitate tagging  
PT of most human genes  
PS Example 4; Page 339; 500pp; English.  
CC The Expressed Sequence Tag was isolated from a human brain cDNA  
CC library as part of a large set of ESTs which can be used as markers  
CC for human genes transcribed in vivo. They can be used to facilitate  
CC tagging of most human genes, for mapping locations of expressed genes  
CC on chromosomes, for individual or forensic identification, for mapping  
CC locations of disease-associated genes, for identification of tissue

CC type, and for prepn. of antisense sequences, probes and constructs.  
CC EST02470 has a "poor" coding probability as evaluated using the  
CC coding-region prediction program CRM. See also Q59041-Q61440.  
SQ Sequence 249 BP; 63 A; 43 C; 42 G; 96 T;

## Query Match 3.48; Score 18; DB 1; Length 249;

Best Local Similarity 100.0%; Pred. No. 11;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 509 ttctgtcaaaaaaa 526  
DB 121 TTCTGTCAAAAAAAA 104

## RESULT 4

V84482  
ID V84482 standard; DNA; 1381 BP.  
AC V84482;  
DT 01-MAR-1999 (first entry)  
DE Human secreted protein gene 72 clone HHGD013.  
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
KW immune system; asthma; lymphocytic disease; brain; lymphoma;  
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
OS Homo sapiens.  
PN W09854963-A2.  
PD 10-DEC-1998.  
PF 04-JUN-1998; U11422.  
PR 18-DEC-1997; US-070923.  
PR 06-JUN-1997; US-048877.  
PR 06-JUN-1997; US-048881.  
PR 06-JUN-1997; US-048884.  
PR 06-JUN-1997; US-048893.  
PR 06-JUN-1997; US-048896.  
PR 06-JUN-1997; US-048899.  
PR 06-JUN-1997; US-048915.  
PR 06-JUN-1997; US-048949.  
PR 06-JUN-1997; US-048964.  
PR 06-JUN-1997; US-048972.  
PR 06-JUN-1997; US-049020.  
PR 05-SEP-1997; US-049375.  
PR 05-SEP-1997; US-057628.  
PR 05-SEP-1997; US-057635.  
PR 05-SEP-1997; US-057644.  
PR 05-SEP-1997; US-057647.  
PR 05-SEP-1997; US-057650.  
PR 05-SEP-1997; US-057661.  
PR 05-SEP-1997; US-057667.  
PR 05-SEP-1997; US-057761.  
PR 05-SEP-1997; US-057764.  
PR 05-SEP-1997; US-057770.  
PR 05-SEP-1997; US-057775.  
PR 05-SEP-1997; US-057778.  
PR 06-JUN-1997; US-048875.  
PR 06-JUN-1997; US-048878.  
PR 06-JUN-1997; US-048882.  
PR 06-JUN-1997; US-048885.  
PR 06-JUN-1997; US-048894.  
PR 06-JUN-1997; US-048897.  
PR 06-JUN-1997; US-048900.  
PR 06-JUN-1997; US-048916.  
PR 06-JUN-1997; US-048962.  
PR 06-JUN-1997; US-048970.  
PR 06-JUN-1997; US-048974.  
PR 06-JUN-1997; US-049373.  
PR 05-SEP-1997; US-057584.  
PR 05-SEP-1997; US-057629.  
PR 05-SEP-1997; US-057642.

05-SEP-1997; US-057645.  
05-SEP-1997; US-057648.  
05-SEP-1997; US-057651.  
05-SEP-1997; US-057662.  
05-SEP-1997; US-057668.  
05-SEP-1997; US-057762.  
05-SEP-1997; US-057765.  
05-SEP-1997; US-057765.  
05-SEP-1997; US-057771.  
05-SEP-1997; US-057776.  
05-SEP-1997; US-057776.  
06-JUN-1997; US-048876.  
06-JUN-1997; US-048880.  
06-JUN-1997; US-048883.  
06-JUN-1997; US-048892.  
06-JUN-1997; US-048895.  
06-JUN-1997; US-048898.  
06-JUN-1997; US-048901.  
06-JUN-1997; US-048917.  
06-JUN-1997; US-048963.  
06-JUN-1997; US-048971.  
06-JUN-1997; US-049019.  
06-JUN-1997; US-049374.  
05-SEP-1997; US-057627.  
05-SEP-1997; US-057634.  
05-SEP-1997; US-057643.  
05-SEP-1997; US-057646.  
05-SEP-1997; US-057649.  
05-SEP-1997; US-057654.  
05-SEP-1997; US-057666.  
05-SEP-1997; US-057760.  
05-SEP-1997; US-057763.  
05-SEP-1997; US-057769.  
05-SEP-1997; US-057774.  
05-SEP-1997; US-057777.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,  
PI Fan P, Feng P, Ferrie AM, Fischer CL, Florence C,  
PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW,  
PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,  
PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;  
DR WPI; 99-059885/05.  
DR P-PSDB; W88605.  
PT New isolated human genes and the secreted polypeptides they encode -  
PT useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders  
PS Claim 4; Page 336-337; 77pp; English.  
CC The invention relates to nucleic acid sequences (W84411 to W84633)  
CC encoding human secreted proteins (W88534 to W88756). The secreted protein  
CC gene sequences are deposited with the ATCC under deposit numbers ATCC  
CC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,  
CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host  
CC cells comprising recombinant vectors containing the nucleic acid  
CC sequences are used for the recombinant production of the secreted  
CC proteins. The polynucleotide and amino acid sequences are useful for are  
CC useful for preventing, treating or ameliorating medical conditions e.g.  
CC by protein or gene therapy. Pathological conditions can be also diagnosed  
CC by determining the amount of the new polypeptides in a sample or by  
CC determining the presence of mutations in the new polynucleotides.  
CC Specific uses are described for each of the polynucleotides, based on  
CC which tissues they are most highly expressed in, and include developing  
CC products for the diagnosis or treatment of cancer, neurodegenerative  
CC disorders, developmental abnormalities and foetal deficiencies, blood  
CC disorders, tumours, leukemias, diseases of the immune system, autoimmune  
CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,  
CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,  
CC restenosis, prostate diseases, obesity, disorders involving osteoclasts  
CC such as osteoporosis, arthritis or malignancies, diseases of testes, lung  
CC or thymus, digestive/endocrine disorders, infections and AIDS. The  
CC polypeptides are also useful for identifying their binding partners.  
CC The present sequence represents a gene encoding a human secreted protein  
CC (see descriptor line for gene number and clone identification).  
SQ Sequence 1381 BP; 328 A; 339 C; 346 G; 365 T;

Query Match 3.4%; Score 18; DB 1; Length 1381;  
Best Local Similarity 100.0%; Pred.No.11;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 509 ttctgtcaaaaaaaaaa 526  
|||||  
DB 1353 TTCGTCAAAAAAAAAA 1370  
RESULT 5  
Q03827  
ID Q03827 standard; DNA; 1713 BP.  
AC Q03827; 1713 BP.  
DT 15-FEB-1993 (first entry)  
DE Pyruvate oxidase (wild-type).  
KW POD; mutation; decarboxylation; assay; ss.  
OS Synthetic.  
PN DE3833601-A.  
PD 05-APR-1990.  
PF 03-OCT-1988; 833601.  
PR 03-OCT-1988; DE-833601.  
PA (BOEF ) BOEHRINGER MANNHEIM GMBH.  
PI Mollering H, Schumacher G;  
DR WPI; 90-108586/15.  
DR P-PSDB; R05793.  
PT New stable, mutated forms of pyruvate oxidase - having specific  
PT aminoacid substitutions, useful as assay reagents, are encoded in  
PT new DNA  
PS Claim 16; Page 6; 10pp; German.  
CC The DNA sequence of wild-type POD, given below, may be mutated so  
CC that at least 178-Pro and/or 425-Ala of the encoded POD are exchanged.  
CC This may be achieved by mutation of nucleotides 532, 533, 534, 1273,  
CC 1274 and/or 1275. Esp. the mutation is of nucleotides 532 and/or 1274  
CC from C to T, resulting in exchange from 178-Pro to Ser and/or 425-Ala  
CC to Val.  
CC The mutated pyruvate oxidase (POD) decarboxylates pyruvate with  
CC formation of H2O2 and is active without addn. of FAD, thiamine  
CC pyrophosphate and divalent metal ions. It is more stable (esp. in  
CC presence of salts and at alkaline pH) than wild-type enzyme, and is  
CC better suited for assay of pyruvate, or pyruvate-generating reactions.  
CC See also Q08597.  
SQ Sequence 1713 BP; 502 A; 315 C; 405 G; 491 T;  
Query Match 3.4%; Score 18; DB 1; Length 1713;  
Best Local Similarity 100.0%; Pred.No.11;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 97 gttataaaattgcggag 114  
|||||  
DB 1039 GTTAATAAAATTCGGAG 1056  
RESULT 6  
Q08597  
ID Q08597 standard; DNA; 1713 BP.  
AC Q08597;  
DT 15-FEB-1993 (first entry)  
DE Pyruvate oxidase (C532T, C1274T).  
KW POD; mutation; decarboxylation; assay; ss.  
OS Synthetic.  
PN DE3833601-A.  
PD 05-APR-1990.  
PF 03-OCT-1988; 833601.  
FT mutation 532 Location/Qualifiers  
FT mutation /note= "C -> T; at least one of wild-type C532  
FT and C1274 is exchanged for T".  
FT mutation 1274  
FT mutation /note= "C -> T; at least one of wild-type C532  
FT and C1274 is exchanged for T".  
PN DE3833601-A.  
PD 05-APR-1990.  
PF 03-OCT-1988; 833601.

PR 03-OCT-1988; DE-833601.  
PA (BOEF) BOEHRINGER MANNHEIM GMBH.  
PI Mollering H, Schumacher G;  
DR WPI; 90-108586/15.  
PT P-PSDB; R09316.  
PT New stable, mutated forms of pyruvate oxidase - having specific  
PT aminoacid substitutions, useful as assay reagents, are encoded in  
PT new DNA  
PS Claim 11; Page 6; 10pp; German.  
CC The DNA sequence of wild-type POD, given in Q03827, may be mutated so  
CC that at least 178-Pro and/or 425-Ala of the encoded POD are exchanged.  
CC This may be achieved by mutation of nucleotides 532, 533, 534, 1273,  
CC 1274 and/or 1275.  
CC For example, the sequence given below comprises mutations of  
CC nucleotides 532 and 1274 from C to T, resulting in exchange from  
CC 178-Pro to Ser and 425-Ala to Val.  
CC The mutated pyruvate oxidase (POD) decarboxylates pyruvate with  
CC formation of H2O2 and is active without addn. of FAD, thiamine  
CC pyrophosphate and divalent metal ions. It is more stable (esp. in  
CC presence of salts and at alkaline pH) than wild-type enzyme, and is  
CC better suited for assay of pyruvate, or pyruvate-generating reactions.  
SQ Sequence 1713 BP; 502 A; 313 C; 405 G; 493 T;

Query Match 3.4%; Score 18; DB 1; Length 1713;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 97 gtaataaaattcgag 114  
|||||  
DB 1039 GTTAATAAAATTCGGAG 1056

RESULT 7  
X07195/c  
ID X07195 standard; DNA; 245 BP.  
AC X07195;  
DT 21-MAY-1999 (first entry)  
DE Human endogenous retrovirus 5' STR.  
KW HERV; IDDMK1.2-22; superantigen; SAG; antigen; IDDM;  
KW insulin-dependent diabetes mellitus; autoimmune disease; diagnosis;  
KW therapy; vaccine; ss.  
OS Human endogenous retrovirus.  
PN W09905527-A2.  
PD 04-FEB-1998; E04926.  
PF 22-JUL-1997; EP-401773.  
PR 23-JUL-1997; EP-112482.  
PA (MEDI-) MEDIGEN SA.  
PI Conrad B, Mach B;  
DR WPI; 99-143118/12.  
PT New isolated human endogenous retrovirus - used to develop products  
PT for the diagnosis, prevention and treatment of autoimmune disease,  
PT particularly insulin dependent diabetes mellitus  
PS Disclosure; Fig 8D; 16pp; English.  
CC This is a retroviral 5' sequence (STR) isolated in the first step  
CC of a 6-step procedure to isolate putative retroviral genomes from  
CC insulin-dependent diabetes mellitus (IDDM) patients. A novel human  
CC endogenous retrovirus (HERV), designated IDDMK1.2-22, has been  
CC identified as the source of superantigen (SAG) activity in IDDM  
CC patients. It is ubiquitous in the human genome but is only  
CC expressed in diabetic individuals. The invention provides  
CC IDDMK1.2-22 nucleic acids (see X07186-91) and encoded polypeptides  
CC (see W97745-48), including env polypeptides responsible for SAG  
CC activity. These nucleic acids and polypeptides can be used in  
CC methods for the diagnosis, treatment and prevention of IDDM.  
CC Expression of an endogenous SAG in IDDM suggests a general model  
CC according to which self SAG-driven and systemic activation of  
CC autoreactive T cells leads to organ-specific autoimmune disease.  
SQ Sequence 245 BP; 49 A; 68 C; 52 G; 76 T;

Query Match 3.2%; Score 17; DB 1; Length 245;

Best Local Similarity 100.0%; Pred. No. 34;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 360 gaataaagacacaaga 376  
|||||  
DB 189 GAAATAAAGACACAAGA 173  
RESULT 8  
X07520/c  
ID X07520 standard; DNA; 245 BP.  
AC X07520;  
DT 08-JUN-1999 (first entry)  
DE Human endogenous retrovirus IDDMK1.2-22 5' end sequence (STR).  
KW IDDM; insulin-dependent diabetes mellitus; endogenous retrovirus;  
KW SAG; superantigen; provirus; autoimmune disease; type 1 diabetes;  
KW diagnosis; short terminal repeat; ss.  
OS Homo sapiens.  
PN EP-893691-A1.  
PD 27-JAN-1999.  
PF 23-JUL-1997; 401773.  
PR 23-JUL-1997; EP-401773.  
PA (MACH/) MACH B F.  
PI Conrad B, Mach B;  
DR WPI; 99-097928/09.  
PT Diagnosing human autoimmune disease by detecting retrovirus with  
PT superantigen activity - new retrovirus associated with type 1  
PT diabetes, its proviral DNA, and related binding agents, transformed cells,  
PT proteins, antibodies and specific binding agents, used for treating  
PT or preventing autoimmune disease  
PS Disclosure; Fig 8D; 92pp; English.  
CC The sequence is that of an insulin-dependent diabetes mellitus  
CC associated human endogenous retrovirus (IDDMK1.2-22) STR.  
CC The retrovirus has Superantigen (SAG) activity. It can be used  
CC as part of a method is specifically used to diagnose type 1 diabetes  
CC mellitus. Modified proteins expressed by the retroviral sequence  
CC (without SAG activity but still able to induce an immune response)  
CC are useful in vaccines to treat or prevent SAG-related autoimmune  
CC disease; nucleic acid sequences encoding (modified) SAG can be used  
CC similarly to treat such diseases. Retroviral-encoded SAG are important  
CC in pathogenesis of autoimmune disease, probably by activating  
CC autoreactive T cells. The method is very specific (it can differentiate  
CC between expressed and non-expressed viral nucleic acids) and can be used  
CC even where the pathogen is an ubiquitous endogenous retrovirus. Blood  
CC or plasma samples can be tested without extensive preparation and  
CC diagnosis can be made before clinical signs are apparent, allowing  
CC early intervention before severe tissue damage has occurred.  
SQ Sequence 245 BP; 49 A; 68 C; 52 G; 76 T;

Query Match 3.2%; Score 17; DB 1; Length 245;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 360 gaataaagacacaaga 376  
|||||  
DB 189 GAAATAAAGACACAAGA 173

RESULT 9  
V69628/c  
ID V69628 standard; DNA; 299 BP.  
AC V69628;  
DT 28-JAN-1999 (first entry)  
DE Human secreted protein gene 18 clone HRLMD77.  
KW Secreted protein; gene therapy; protein therapy; diagnosis; treatment;  
KW central nervous system; CNS; immune system; cancer; trauma; liver;  
KW reproductive disorder; congenital malformation; degenerative disease;  
KW inflammatory disease; neoplasia; metabolic disorder; testis; placenta;  
KW brain; T cell; spleen; lung; heart; rhabdomyosarcoma; endocrine system;  
KW endocrinopathy; endocrine polyglandular syndrome; endocrinoma; sepsis;  
KW endocrine ophthalmopathy; osteoclastoma; bacterial infection; bone; ds.  
OS Homo sapiens.

```

FH Key          Location/Qualifiers
FT CDS          153..299
FT              /*tag= a
FT              /product= "secreted protein"
FT              /note= "the stop codon is not indicated"
FT sig_peptide  153..221
FT              /*tag= b
FT mat_peptide  222..299
FT              /*tag= c
FN WO9845712-A2.
PD 15-OCT-1998.
PF 07-APR-1998; U06801.
PR 30-MAY-1997; US-048184.
PR 08-APR-1997; US-042726.
PR 08-APR-1997; US-042727.
PR 08-APR-1997; US-042728.
PR 08-APR-1997; US-042754.
PR 08-APR-1997; US-042825.
PR 30-MAY-1997; US-048068.
PR 30-MAY-1997; US-048070.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Feng P, Ni J, Rosen CA, Ruben SM, Yu G;
DR WPI; 98-594496/50.
DR P-PSDB; W83948.
PT New isolated human genes and secreted polypeptide(s) they encode -
PT useful for the diagnosis and treatment of e.g. cancers, CNS
PT disorders, immune system disorders, inflammatory disease and
PT bacterial infections
PS Claim 4; Page 115; 142pp; English.
CC This sequence represents a nucleic acid molecule designated Gene 18 from
CC the human cDNA clone HRLMD77 (deposited as clone ATCC 97955 and ATCC
CC 209074) which encodes a secreted human protein. This sequence shares
CC homology with UFO oncoprotein and is useful for the treatment of cancer.
CC The invention relates to 20 novel genes and their fragments (V69611 to
CC V69630) and corresponding secreted proteins (W83931 to W83950) which are
CC useful for preventing, treating or ameliorating medical conditions e.g.
CC by protein of gene therapy. Also pathological conditions can be diagnosed
CC by determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the polynucleotides. Specific
CC uses are based on which tissues they are most highly expressed in, and
CC include developing products for the diagnosis or treatment of central
CC nervous system (CNS) and immune system diseases, reproductive disorders,
CC cancers, congenital malformations, degenerative diseases, trauma, in testes,
CC inflammatory disease, neoplasia, metabolic disorders, diseases in spleen,
CC placenta, liver, brain and activated T cells, spleen diseases, lung
CC diseases, heart diseases, rhabdomyosarcoma and disorders of the endocrine
CC system or other endocrinopathies, e.g. endocrine polyglandular syndrome,
CC endocrinoma, and endocrine ophthalmopathy, osteoclastoma and other bone
CC remodelling disorders, bacterial infections and sepsis. The polypeptides
CC are also useful for identifying their binding partners.
SQ Sequence 299 BP; 79 A; 78 C; 52 G; 90 T;

Query Match          3.2%; Score 17; DB 1; Length 299;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 510 tctgtcaaaaaa 526
Db 33 TCTGTCAAAAAA 17

RESULT 10
X23242/c
ID X23242 standard; DNA; 312 BP.
AC X23242;
DT 11-JUN-1999 (first entry)
DE mak TTC8 variable region light chain DNA fragment.
KW Monoclonal antibody; hypervariable domain; light chain; neutralise;
KW Clostridium difficile; epitope; ligand domain; immunotherapy; disease;
KW translocation domain; catalytic domain; humanised antibody; enterotoxin;
KW toxin A; toxin B; pseudomembranous colitis; ss.
OS Synthetic.

```

```

PN DE19739685-A1.
PD 11-MAR-1999.
PF 10-SEP-1997; 039685.
PR 10-SEP-1997; DE-039685.
PA (VEIC/) VON EICHEL-STREIBER C.
PI Moos M, Von Eichel-Streiber C;
DR WPI; 99-182094/16.
DR P-PSDB; W93481.
PT Monoclonal antibodies specific for Clostridium difficile toxins -
PT especially humanised antibodies for treating pseudomembranous
PT colitis
PS Disclosure; Page 13; 14pp; German.
CC This invention describes a novel monoclonal antibody that is directed
CC against a Clostridium difficile toxin and recognises and neutralises
CC an epitope in the ligand domain, translocation domain or catalytic
CC domain of the toxin. Humanised antibodies are described which are
CC expressed in plants and can be used for immunotherapy of diseases caused
CC by Clostridium difficile enterotoxin (toxin A) or cytotoxin (toxin B),
CC especially pseudomembranous colitis.
SQ Sequence 312 BP; 81 A; 82 C; 73 G; 76 T;

Query Match          3.2%; Score 17; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 149 tggacatgaatttggga 165
Db 31 TGGACATGAATTTTGGGA 15

RESULT 11
V82857/c
ID V82857 standard; cDNA to mRNA; 318 BP.
AC V82857;
DT 25-FEB-1999 (first entry)
DE Nucleic acid encoding light chain V region of an anti-E2 antibody.
KW Light chain V region; antibody; estradiol; E2; chimeric antibody; ds.
OS Mus sp.
PN J10309198-A.
PD 24-NOV-1998.
PF 12-MAY-1997; 120596.
PR 12-MAY-1997; JP-120596.
PA (TOIJ) TOSOH CORP.
DR WPI; 99-063679/06.
DR P-PSDB; W85537.
PT Novel gene coding H chain V region or L chain V region of antibody
PT recognising oestradiol - and preparation of a antibody using the
PT gene
PS Claim 2; Page 6; 7pp; Japanese.
CC The present sequence encodes the light chain V region of an antibody
CC recognising estradiol (E2). The sequence can be used to prepare a
CC chimeric antibody.
SQ Sequence 318 BP; 88 A; 83 C; 77 G; 70 T;

Query Match          3.2%; Score 17; DB 1; Length 318;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 149 tggacatgaatttggga 165
Db 37 TGGACATGAATTTTGGGA 21

RESULT 12
V86315
ID V86315 standard; cDNA; 330 BP.
AC V86315;
DT 27-APR-1999 (first entry)
DE EST clone AM235.
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibitor; tumour invasion suppressor; EST; human;

```

KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
 OS Homo sapiens.  
 PN WO9845435-A2.  
 PD 15-OCT-1998.  
 PF 10-APR-1998; U06954.  
 PR 10-APR-1997; US-835913.  
 PA (GEMV) GENETICS INST INC.  
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
 PI Racie LA, Spaulding V, Treacy M;  
 DR WPI; 99-070076/06.  
 PT New polynucleotides encoding human secreted proteins - derived from  
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PT ovary, pituitary, retina and colon cDNA libraries  
 PS Claim 1; Page 193; 633pp; English.  
 CC This sequence represents an expressed sequence tag (EST), and is a  
 CC polynucleotide of the invention. The polynucleotides of the invention are  
 CC all secreted EST sequences isolated from a variety of human tissue  
 CC sources. The EST sequences and proteins encoded by them are predicted to  
 CC have useful biological activities which would make them suitable for  
 CC treating, preventing or ameliorating medical conditions in humans and  
 CC animals, although no supporting data is given. Suggested activities  
 CC include nutritional activity, immune stimulating or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cacharin/tumour invasion suppressor activity, tumour inhibition  
 CC therapy. The EST sequences are also stated to be useful for gene  
 SQ Sequence 330 BP; 117 A; 63 C; 47 G; 103 T;

Query Match 3.2%; Score 17; DB 1; Length 330;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 469 cctggaaattactgtgta 485  
 Db 206 CCGTGAATTAATCTGTA 222  
 |||||

RESULT 13  
 T12399/c  
 ID T12399 standard; cDNA: 472 BP.  
 AC T12399;  
 DE 3'-flanking sequence of alpha-lactalbumin gene.  
 KW Human; alpha-lactalbumin; 3'-flanking sequence; milk; vector;  
 KW gene transfer; tissue-specific gene expression; transgenic animal;  
 KW cattle; mammary; plasmid pBBHA; plasmid pOBHA; plasmid pBAHA;  
 KW plasmid pBova-A; plasmid pBova-B; plasmid pHA1; plasmid pHA2;  
 KW infant formula; pharmaceutical; ds.  
 OS Homo sapiens.  
 PN WO9602640-A1.  
 PD 01-FEB-1996.  
 PF 12-JUL-1995; G01651.  
 PR 13-JUL-1994; WO-G01514.  
 PR 15-DEC-1994; GB-025326.  
 PR 31-JAN-1995; US-381691.  
 PR 25-FEB-1995; GB-003822.  
 PA (PPLT-) PPL THERAPEUTICS SCOTLAND LTD.  
 PI Cooper J, Schnieke A;  
 DR WPI; 96-105903/11.  
 PT Alpha-lactalbumin gene constructs - esp. for expressing human  
 PT alpha-lactalbumin in milk of transgenic animals for use in human  
 PT infant nutrition  
 PS Claim 12; Page 51; 77pp; English.  
 CC The sequence is a fragment (nucleotides 3481-3952) of the  
 CC 3'-flanking sequence of a human alpha-lactalbumin gene, encoding a  
 CC protein found in human milk. The sequence has been isolated by  
 CC polymerase chain reaction amplification of 2 fragments of human  
 CC genomic DNA, and use of these as probes to screen a phage lambda  
 CC genomic library. Other 3'-flanking sequences are given in

CC T12395-98, and a 5'-flanking sequence is given in T12400. Inclusion  
 CC of these sequences in a vector results in enhanced expression of a  
 CC transgene (encoding alpha-lactalbumin or a heterologous protein) in  
 CC the mammary gland. The vector (e.g. plasmid pBBHA, plasmid pOBHA,  
 CC plasmid pBAHA, plasmid pBova-A, plasmid pBova-B, plasmid pHA1 or  
 CC plasmid pHA2 (NCIMB 40709)) may be used to express recombinant  
 CC proteins in the milk of transgenic animals (e.g. cattle), for use  
 CC e.g. in an infant formula, or in pharmaceutical protein production.  
 SQ Sequence 472 BP; 104 A; 110 C; 103 G; 155 T;

Query Match 3.2%; Score 17; DB 1; Length 472;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 510 tctgtcaaaaaaaaaa 526  
 Db 378 TCTGTCAAAAAAAAAA 362  
 |||||

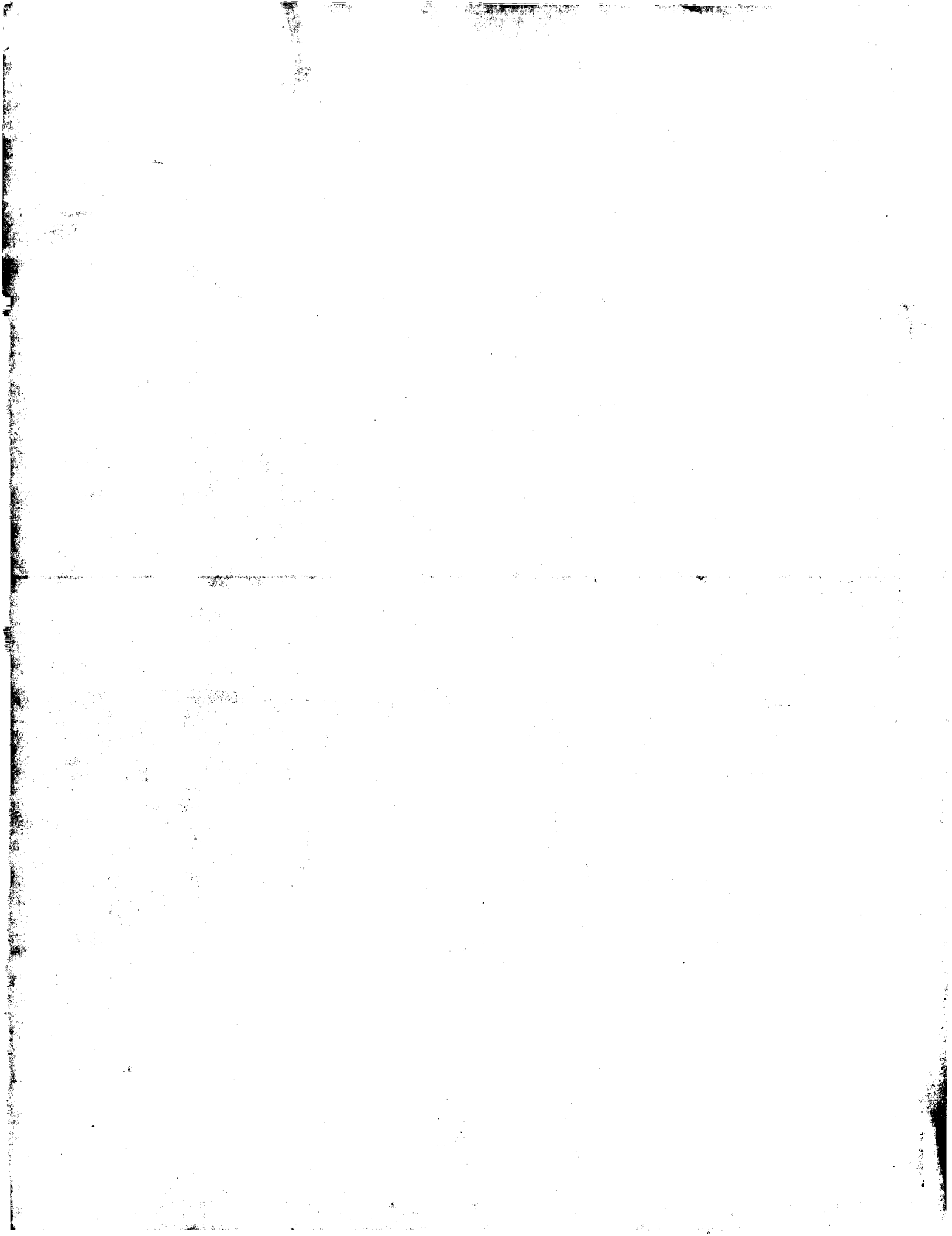
RESULT 14  
 T47733/c  
 ID T47733 standard; cDNA: 723 BP.  
 AC T47733;  
 DE 23-SEP-1997 (first entry)  
 DE Anti-CD19 antibody SJ25c1 single chain variable region cDNA.  
 KW CD19; antibody; variable region; B43; SJ25c1; BLV3; targeting;  
 KW cytotoxin conjugate; cancer treatment; leukaemia; B cell lymphoma;  
 KW variable; heavy; light chain; ds.  
 OS Synthetic.  
 FH Key  
 FH Location/Qualifiers  
 Key 1..723  
 FT /\*tag= a  
 FT /note= "no start or stop codon given"  
 FT mat\_peptide 1..345  
 FT /\*tag= b  
 FT /note= "encodes heavy chain"  
 FT misc\_feature 346..390  
 FT /\*tag= c  
 FT /note= "encodes linker"  
 FT mat\_peptide 391..741  
 FT /\*tag= d  
 FT /note= "encodes light chain"  
 FT WO9636360-A1.  
 PN 21-NOV-1996.  
 PD 15-MAY-1996; U06941.  
 PR 17-MAY-1995; US-443408.  
 PA (MINU) UNIV MINNESOTA.  
 PI Bejcek BE, Kersey JH, Uckun FM, Wang D;  
 DR WPI; 97-042642/04.  
 PT DNA encoding a single chain variable region polypeptide which binds  
 PT CD19 - used in the treatment of leukaemia and B-cell lymphoma  
 PS Claim 3; Page 102-103; 143pp; English.  
 CC T47733 encodes a single chain variable region polypeptide  
 CC (heavy chain-linker-light chain) of anti-CD19 antibody SJ25c1. Single  
 CC chain polypeptides were conjugated to at least one cytotoxic agent and  
 CC used to target the agent(s) to cancerous cells expressing CD19. The  
 CC immunoconjugates can be used treatment of cancer, especially leukaemia  
 CC and B-cell lymphoma.  
 SQ Sequence 723 BP; 184 A; 177 C; 200 G; 162 T;

Query Match 3.2%; Score 17; DB 1; Length 723;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 tggacatgaatttggg 165  
 Db 427 TGGACATGAATTTGGA 411  
 |||||

RESULT 15  
 V11399/c







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:01:59 ; Search time 181.17 Seconds  
(without alignments)  
347.598 Million cell updates/sec

Title: US-09-215-435-76  
Perfect score: 526  
Sequence: 1 ctgctgtgtgtgtgtgcac.....cttctgtcaaaaaaaaaa 526

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 214294 seqs, 59861574 residues

Word size : 0

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/5C.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/5D.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/6.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PTUS9.COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	3.2	1196	5	US-07-959-509-4
2	17	3.2	1505	4	US-08-909-965C-13
3	17	3.2	2188	1	US-07-882-925A-4
4	17	3.2	2188	1	US-08-184-012C-4
5	17	3.2	3348	1	US-08-222-616-35
6	17	3.2	3348	6	PCT-US95-04228-35
7	17	3.2	3952	3	US-08-381-691-16
8	17	3.2	8174	1	US-07-914-281-5
9	17	3.2	8174	1	US-08-393-246-5
10	17	3.2	8174	2	US-08-525-058A-5
11	17	3.2	8174	4	US-08-696-731-5
12	17	3.2	8174	6	PCT-US91-00899-3
13	17	3.2	246240	3	US-08-724-394A-20
14	17	3.2	246240	3	US-08-724-394A-21
15	17	3.2	246240	3	US-08-724-394A-22
16	16	3.0	61	1	US-08-384-490-14
17	16	3.0	61	1	US-08-459-383-14
18	16	3.0	961	6	PCT-US91-08177-14
19	16	3.0	965	2	US-08-257-073-6
20	16	3.0	1415	1	US-08-236-918A-7
21	16	3.0	1416	2	US-08-535-237-1
22	16	3.0	1494	4	US-08-583-562B-11
23	16	3.0	1494	4	US-08-779-113-11
24	16	3.0	1666	1	US-08-076-090-1
25	16	3.0	1666	6	PCT-US94-06661-1
26	16	3.0	1833	7	5166195-1

c 27	16	3.0	2500	2	US-08-550-715-10	Sequence 10, Appl
c 28	16	3.0	2574	4	US-08-583-562B-1	Sequence 1, Appl
c 29	16	3.0	2574	4	US-08-779-113-1	Sequence 1, Appl
c 30	16	3.0	2652	3	US-08-366-547-1	Sequence 1, Appl
c 31	16	3.0	4146	1	US-08-261-822A-15	Sequence 15, Appl
c 32	16	3.0	4146	6	PCT-US95-07744A-15	Sequence 15, Appl
c 33	16	3.0	6060	6	PCT-US96-09430-7	Sequence 7, Appl
c 34	16	3.0	7323	6	PCT-US91-08177-1	Sequence 1, Appl
c 35	16	3.0	8920	3	US-08-446-855A-1	Sequence 1, Appl
c 36	15	2.9	38	4	US-08-415-823-6	Sequence 6, Appl
c 37	15	2.9	38	4	US-09-086-662-6	Sequence 23, Appl
c 38	15	2.9	459	2	US-08-468-347-23	Sequence 23, Appl
c 39	15	2.9	459	2	US-08-226-264-25	Sequence 25, Appl
c 40	15	2.9	459	2	US-08-467-389-23	Sequence 23, Appl
c 41	15	2.9	459	3	US-08-779-379-23	Sequence 23, Appl
c 42	15	2.9	469	3	US-08-469-219-23	Sequence 1, Appl
c 43	15	2.9	729	2	US-08-447-010-1	Sequence 1, Appl
c 44	15	2.9	857	4	US-08-627-610-5	Sequence 5, Appl
c 45	15	2.9	1002	2	US-08-463-090B-3	Sequence 3, Appl

## ALIGNMENTS

RESULT 1  
US-07-959-509-4  
; Sequence 4, Application US/07959509  
; Patent No. 6001560  
; GENERAL INFORMATION:  
; APPLICANT: Lonial, Herinder  
; APPLICANT: Narula, Satwant  
; APPLICANT: Zavodny, Paul  
; TITLE OF INVENTION: Human Gamma Interferon Antagonist/Agonist Screen  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Schering-Plough Corporation  
; STREET: One Giralda Farms  
; CITY: Madison  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07940  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Macintosh 6.0.5  
; SOFTWARE: Microsoft Word 4.00B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07959,509  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/616,621  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dulak, No. 6001560man C.  
; REGISTRATION NUMBER: 31,608  
; REFERENCE/DOCKET NUMBER: JBO166  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201 822 7375  
; TELEFAX: 201 822 7039  
; TELEX: 219165  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1196 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-07-959-509-4

Query Match 3.28; Score 17; DB 5; Length 1196;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```
; Patent No. 5606029
; GENERAL INFORMATION:
; APPLICANT: Degen, Sandra J. F.
; TITLE OF INVENTION: Gene for a growth factor and its cDNA and
; TITLE OF INVENTION: protein
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gregory Lunn
; STREET: Wood, Herron & Evans, 2700 Carew Tower
; CITY: Cincinnati
; STATE: Ohio
; COUNTRY: USA
; ZIP: 45202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.5.2
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/184.012C
; FILING DATE: 1/18/94
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Lunn, Gregory
; REGISTRATION NUMBER: 29,945
; REFERENCE/DOCKET NUMBER: CMC 57
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 241-2324
; TELEFAX: (513) 421-7269
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2188 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: mouse
; STRAIN: C57BL/6
; DEVELOPMENTAL STAGE: adult
; TISSUE TYPE: liver
; IMMEDIATE SOURCE:
; LIBRARY: cDNA
; CLONE: ML5-2
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: mouse 9, Hgfl locus
; MAP POSITION: Trf-Gnai-2-Hgfl-Cck
; FEATURE:
; IDENTIFICATION METHOD: experimental
; PUBLICATION INFORMATION:
; RELEVANT RESIDUES IN SEQ ID NO: 4: 1 TO 2188
US-08-184-012C-4
```

```
Query Match 3.2%; Score 17; DB 1; Length 2188;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 510 tctgtcaaaaaa 526
| | | | | | | | | |
Db 2167 TCTGTCAAAAAA 2183
```

```
RESULT 5
US-08-222-616-35
; Sequence 35, Application US/08222616
; Patent No. 5635177
; GENERAL INFORMATION:
; APPLICANT: Bennett, Brian D.
; APPLICANT: Goeddel, David
; APPLICANT: Lee, James M.
; APPLICANT: Matthews, William
```

```
; APPLICANT: Tsai, Siao Ping
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST
; TITLE OF INVENTION: ANTIBODIES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222.616
; FILING DATE: 4-APR-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/00586
; FILING DATE: 22-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/826935
; FILING DATE: 22-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 821P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3348 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-222-616-35
```

```
Query Match 3.2%; Score 17; DB 1; Length 3348;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 141 ctgcaaatggacatga 157
| | | | | | | | | |
Db 360 CTGCAATGGACATGA 376
```

```
RESULT 6
PCI-US95-04228-35
; Sequence 35, Application PC/TUS9504228
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Bennett, Brian D.
; APPLICANT: Goeddel, David
; APPLICANT: Lee, James M.
; APPLICANT: Matthews, William
; APPLICANT: Tsai, Siao Ping
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
```



STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/393,246  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/220,433  
FILING DATE: 30-MAR-1994  
APPLICATION NUMBER: US 07/914,281  
FILING DATE: 20-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Lavalleye, Jean-Paul M. P.  
REGISTRATION NUMBER: 31,451  
REFERENCE/DOCKET NUMBER: 2363-060-55  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)521-4500  
TELEFAX: (703)486-2347  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8174 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: DNA (genomic)  
MOLECULE TYPE: NO  
ANTI-SENSE: NO  
US-08-393-246-5

Query Match 3.2% Score 17; DB 1; Length 8174;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 360 gaaataaagacacacaaga 376  
|||||  
Db 2381 GAAATAAAGACACACAGA 2365  
RESULT 10  
US-08-525-058A-5/c  
Sequence 5, Application US/08525058A  
Patent No. 5770420  
GENERAL INFORMATION:  
APPLICANT: LOWE, JOHN B.  
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS  
OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,  
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION  
OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.  
STREET: 1755 Jefferson Davis Highway, Fourth Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/525,058A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M. P.  
REGISTRATION NUMBER: 31,451  
REFERENCE/DOCKET NUMBER: 2363-060-55  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)521-4500  
TELEFAX: (703)486-2347  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8174 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ANTI-SENSE: NO  
US-08-525-058A-5  
Query Match 3.2% Score 17; DB 2; Length 8174;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 360 gaaataaagacacacaaga 376  
|||||  
Db 2381 GAAATAAAGACACACAGA 2365  
RESULT 11  
US-08-696-731-5/c  
Sequence 5, Application US/08696731  
Patent No. 5955347  
GENERAL INFORMATION:  
APPLICANT: LOWE, JOHN B.  
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS  
OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,  
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION  
OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
STREET: 1755 Jefferson Davis Highway, Fourth Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/696,731  
FILING DATE: 14-AUG-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/393,246  
FILING DATE:  
APPLICATION NUMBER: US 08/220,433  
FILING DATE: 30-MAR-1994  
APPLICATION NUMBER: US 07/914,281  
FILING DATE: 20-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Lavalleye, Jean-Paul M. P.  
REGISTRATION NUMBER: 31,451  
REFERENCE/DOCKET NUMBER: 2363-060-55  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)521-4500  
TELEFAX: (703)486-2347  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8174 base pairs

;; TYPE: nucleic acid  
;; STRANDEDNESS: unknown  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: DNA (genomic)  
;; ANTI-SENSE: NO  
US-08-696-731-5

Query Match 3.2%; Score 17; DB 4; Length 8174;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 360 gaaataaagacacaaga 376  
|||||  
Db 2381 GAAATAAGACACAAGA 2365

## RESULT 12

PCT-US91-00899-3/c

; Sequence 3, Application PC/TUS9100899

; GENERAL INFORMATION:

; APPLICANT: Lowe, John B.

; TITLE OF INVENTION: Method and Products For the Synthesis of

; TITLE OF INVENTION: Oligosaccharide Structures on Glycoproteins, Glycolipids,

; TITLE OF INVENTION: or as Free Molecules, and For the Isolation of Cloned

; TITLE OF INVENTION: Genetic Sequences That Determine These Structures

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER &amp; NEUSTADT,

; STREET: P.C.

; STREET: 1755 Jefferson Davis Highway, Suite 400

; CITY: Arlington

; STATE: Virginia

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION NUMBER: PCT/US91/00899

; FILING DATE: 19910214

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Lavalleye Ph.D., Jean-Paul

; REGISTRATION NUMBER: 31,451

; REFERENCE/DOCKET NUMBER: 2363-021-55 PCT

; TELEPHONE: (703)521-5940

; TELEFAX: (703)486-2347

; TELEX: 248855 OPAT UR

; INFORMATION FOR SEQ ID NO: 3:

; LENGTH: 8174 base pairs

; TYPE: NUCLEIC ACID

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; MOLECULE TYPE: CDNA

; HYPOTHETICAL: NO

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; TISSUE TYPE: Blood

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 4686..5780

; OTHER INFORMATION: /label= mat\_peptide

PCT-US91-00899-3

Query Match 3.2%; Score 17; DB 6; Length 8174;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 360 gaaataaagacacaaga 376  
|||||  
Db 2381 GAAATAAGACACAAGA 2365

## RESULT 13

US-08-724-394A-20

; Sequence 20, Application US/08724394A

; Patent No. 5872237

; GENERAL INFORMATION:

; APPLICANT: Feder, John N.

; APPLICANT: Kronmal, Gregory S.

; APPLICANT: Lauer, Peter M.

; APPLICANT: Ruddy, David A.

; APPLICANT: Thomas, Winston

; APPLICANT: Tsuchihashi, Zenta

; APPLICANT: Wolff, Roger K.

; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP

; STREET: Two Embarcadero Center, 8th Floor

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/724,394A

; FILING DATE: 01-OCT-1996

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Fitts, Renee A.

; REGISTRATION NUMBER: 35,136

; REFERENCE/DOCKET NUMBER: 017957-000100

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-576-0200

; TELEFAX: 415-576-0300

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 246240 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: CDNA

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 1..246240

; OTHER INFORMATION: /note= "HLA-H.CONTIG"

US-08-724-394A-20

Query Match 3.2%; Score 17; DB 3; Length 246240;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 510 tctgtcaaaaaaaaaa 526  
|||||  
Db 189743 TCTGTCAAAAAAAAAA 189759

## RESULT 14

US-08-724-394A-21

; Sequence 21, Application US/08724394A

; Patent No. 5872237

; GENERAL INFORMATION:

; APPLICANT: Feder, John N.

; APPLICANT: Kronmal, Gregory S.

APPLICANT: Lauer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchihashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
TITLE OF INVENTION: Sequences and Antibodies Thereto  
NUMBER OF SEQUENCES: 31

## CORRESPONDENCE ADDRESS:

ADDRESS: TOWNSEND and TOWNSEND and CREW LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94111-3834

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/724.394A

FILING DATE: 01-OCT-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Fitts, Renee A.

REGISTRATION NUMBER: 35,136

REFERENCE/DOCKET NUMBER: 017957-000100

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-576-0200

TELEFAX: 415-576-0300

INFORMATION FOR SEQ ID NO: 21:

## SEQUENCE CHARACTERISTICS:

LENGTH: 246240 base pairs

TYPE: nucleic acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: CDNA

## FEATURE:

NAME/KEY: misc\_feature

LOCATION: 1..246240

OTHER INFORMATION: /note= "HLA-H.CONTIG"

US-08-724-394A-21

Query Match 3.2%; Score 17; DB 3; Length 246240;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 510 tctgtcaaaaaaaaaaaaa 526

Db 189743 TCTGTCAAAAAAAAAAAAA 189759

## RESULT 15

US-08-724-394A-22

; Sequence 22, Application US/08724394A

; Patent No. 5872237

## GENERAL INFORMATION:

APPLICANT: Feder, John N.

APPLICANT: Kronmal, Gregory S.

APPLICANT: Lauer, Peter M.

APPLICANT: Ruddy, David A.

APPLICANT: Thomas, Winston

APPLICANT: Tsuchihashi, Zenta

APPLICANT: Wolff, Roger K.

TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el

TITLE OF INVENTION: Sequences and Antibodies Thereto

NUMBER OF SEQUENCES: 31

## CORRESPONDENCE ADDRESS:

ADDRESS: TOWNSEND and TOWNSEND and CREW LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/724.394A

FILING DATE: 01-OCT-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Fitts, Renee A.

REGISTRATION NUMBER: 35,136

REFERENCE/DOCKET NUMBER: 017957-000100

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-576-0200

TELEFAX: 415-576-0300

INFORMATION FOR SEQ ID NO: 22:

## SEQUENCE CHARACTERISTICS:

LENGTH: 246240 base pairs

TYPE: nucleic acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: CDNA

## FEATURE:

NAME/KEY: misc\_feature

LOCATION: 1..246240

OTHER INFORMATION: /note= "HLA-H.CONTIG"

US-08-724-394A-22

Query Match 3.2%; Score 17; DB 3; Length 246240;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 510 tctgtcaaaaaaaaaaaaa 526

Db 189743 TCTGTCAAAAAAAAAAAAA 189759

Search completed: May 1, 2000, 16:04:38

Job time: 18343 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 19:59:35 ; Search time 4088.29 Seconds  
(without alignments)  
-441.204 Million cell updates/sec

Title: US-09-215-435-76  
Perfect score: 526  
Sequence: 1 ctgcctgctgctgtgac.....ctttctgtcaaaaaaaaaa 526

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 5265378 seqs, -1714608768 residues

Word size : 0

Total number of hits satisfying chosen parameters: 10530736

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Pending Patents\_NA.\*

1: /cgn2\_6/ptodata/1/pna/PCTUS9\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US081A\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US081B\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US081C\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US081D\_COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US082A\_COMB.seq.\*  
9: /cgn2\_6/ptodata/1/pna/US082B\_COMB.seq.\*  
10: /cgn2\_6/ptodata/1/pna/US082C\_COMB.seq.\*  
11: /cgn2\_6/ptodata/1/pna/US083A\_COMB.seq.\*  
12: /cgn2\_6/ptodata/1/pna/US083B\_COMB.seq.\*  
13: /cgn2\_6/ptodata/1/pna/US084A\_COMB.seq.\*  
14: /cgn2\_6/ptodata/1/pna/US084B\_COMB.seq.\*  
15: /cgn2\_6/ptodata/1/pna/US084C\_COMB.seq.\*  
16: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*  
17: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*  
18: /cgn2\_6/ptodata/1/pna/US087A\_COMB.seq.\*  
19: /cgn2\_6/ptodata/1/pna/US087B\_COMB.seq.\*  
20: /cgn2\_6/ptodata/1/pna/US087C\_COMB.seq.\*  
21: /cgn2\_6/ptodata/1/pna/US088A\_COMB.seq.\*  
22: /cgn2\_6/ptodata/1/pna/US088B\_COMB.seq.\*  
23: /cgn2\_6/ptodata/1/pna/US088C\_COMB.seq.\*  
24: /cgn2\_6/ptodata/1/pna/US088D\_COMB.seq.\*  
25: /cgn2\_6/ptodata/1/pna/US089A\_COMB.seq.\*  
26: /cgn2\_6/ptodata/1/pna/US089B\_COMB.seq.\*  
27: /cgn2\_6/ptodata/1/pna/US089C\_COMB.seq.\*  
28: /cgn2\_6/ptodata/1/pna/US089D\_COMB.seq.\*  
29: /cgn2\_6/ptodata/1/pna/US089E\_COMB.seq.\*  
30: /cgn2\_6/ptodata/1/pna/US089F\_COMB.seq.\*  
31: /cgn2\_6/ptodata/1/pna/US089G\_COMB.seq.\*  
32: /cgn2\_6/ptodata/1/pna/US089H\_COMB.seq.\*  
33: /cgn2\_6/ptodata/1/pna/US089I\_COMB.seq.\*  
34: /cgn2\_6/ptodata/1/pna/US090A\_COMB.seq.\*  
35: /cgn2\_6/ptodata/1/pna/US090B\_COMB.seq.\*  
36: /cgn2\_6/ptodata/1/pna/US090C\_COMB.seq.\*  
37: /cgn2\_6/ptodata/1/pna/US090D\_COMB.seq.\*  
38: /cgn2\_6/ptodata/1/pna/US090E\_COMB.seq.\*  
39: /cgn2\_6/ptodata/1/pna/US091A\_COMB.seq.\*  
40: /cgn2\_6/ptodata/1/pna/US091B\_COMB.seq.\*  
41: /cgn2\_6/ptodata/1/pna/US091C\_COMB.seq.\*  
42: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq.\*  
43: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq.\*

44: /cgn2\_6/ptodata/1/pna/US092C\_COMB.seq.\*  
45: /cgn2\_6/ptodata/1/pna/US092D\_COMB.seq.\*  
46: /cgn2\_6/ptodata/1/pna/US092E\_COMB.seq.\*  
47: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq.\*  
48: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq.\*  
49: /cgn2\_6/ptodata/1/pna/US093C\_COMB.seq.\*  
50: /cgn2\_6/ptodata/1/pna/US093D\_COMB.seq.\*  
51: /cgn2\_6/ptodata/1/pna/US093E\_COMB.seq.\*  
52: /cgn2\_6/ptodata/1/pna/US093F\_COMB.seq.\*  
53: /cgn2\_6/ptodata/1/pna/US094A\_COMB.seq.\*  
54: /cgn2\_6/ptodata/1/pna/US094B\_COMB.seq.\*  
55: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq.\*  
56: /cgn2\_6/ptodata/1/pna/US6001A\_COMB.seq.\*  
57: /cgn2\_6/ptodata/1/pna/US6001B\_COMB.seq.\*  
58: /cgn2\_6/ptodata/1/pna/US6002A\_COMB.seq.\*  
59: /cgn2\_6/ptodata/1/pna/US6002B\_COMB.seq.\*  
60: /cgn2\_6/ptodata/1/pna/US6002C\_COMB.seq.\*  
61: /cgn2\_6/ptodata/1/pna/US6003A\_COMB.seq.\*  
62: /cgn2\_6/ptodata/1/pna/US6003B\_COMB.seq.\*  
63: /cgn2\_6/ptodata/1/pna/US6004A\_COMB.seq.\*  
64: /cgn2\_6/ptodata/1/pna/US6004B\_COMB.seq.\*  
65: /cgn2\_6/ptodata/1/pna/US6004C\_COMB.seq.\*  
66: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq.\*  
67: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq.\*  
68: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq.\*  
69: /cgn2\_6/ptodata/1/pna/US6008A\_COMB.seq.\*  
70: /cgn2\_6/ptodata/1/pna/US6008B\_COMB.seq.\*  
71: /cgn2\_6/ptodata/1/pna/US6008C\_COMB.seq.\*  
72: /cgn2\_6/ptodata/1/pna/US6009A\_COMB.seq.\*  
73: /cgn2\_6/ptodata/1/pna/US6009B\_COMB.seq.\*  
74: /cgn2\_6/ptodata/1/pna/US6010A\_COMB.seq.\*  
75: /cgn2\_6/ptodata/1/pna/US6010B\_COMB.seq.\*  
76: /cgn2\_6/ptodata/1/pna/US6011A\_COMB.seq.\*  
77: /cgn2\_6/ptodata/1/pna/US6011B\_COMB.seq.\*  
78: /cgn2\_6/ptodata/1/pna/US6012A\_COMB.seq.\*  
79: /cgn2\_6/ptodata/1/pna/US6012B\_COMB.seq.\*  
80: /cgn2\_6/ptodata/1/pna/US6013A\_COMB.seq.\*  
81: /cgn2\_6/ptodata/1/pna/US6013B\_COMB.seq.\*  
82: /cgn2\_6/ptodata/1/pna/US6014A\_COMB.seq.\*  
83: /cgn2\_6/ptodata/1/pna/US6014B\_COMB.seq.\*  
84: /cgn2\_6/ptodata/1/pna/US6014C\_COMB.seq.\*  
85: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq.\*  
86: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq.\*  
87: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq.\*  
88: /cgn2\_6/ptodata/2/pna/PCT\_NEW\_COMB.seq.\*  
89: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB.seq.\*  
90: /cgn2\_6/ptodata/2/pna/US07\_NEW\_COMB.seq.\*  
91: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq.\*  
92: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*  
93: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	526	100.0	526	42	US-09-215-435-76
2	281	53.4	539	42	US-09-215-435-303
3	281	53.4	539	67	US-60-069-957-136
4	237	45.1	271	25	US-08-905-279-148
5	21	4.0	762	43	US-09-248-796-9251
6	21	4.0	762	73	US-60-096-409-9251
7	21	4.0	1101	43	US-09-248-796-2842
8	21	4.0	1101	73	US-60-096-409-2842
9	20	3.8	378	84	US-60-145-485-7430
10	20	3.8	383	79	US-60-128-437-754
11	20	3.8	499	86	US-60-168-611-752
12	19	3.6	275	36	US-09-041-720-3385
13	19	3.6	303	22	US-08-822-285-583

```

14: 19 3.6 303 56 US-60-013-696-583
15 19 3.6 312 84 US-60-145-485-6596
16 19 3.6 333 65 US-60-048-722-829
17 19 3.6 372 87 US-60-171-432-9978
18 19 3.6 423 84 US-60-145-485-8101
19 19 3.6 450 53 US-09-431-517-2025
20 19 3.6 637 86 US-60-160-189-458
21 18 3.4 69 92 US-09-215-179-14446
22 18 3.4 188 10 US-08-270-638A-978
23 18 3.4 188 13 US-08-412-033-2335
24 18 3.4 188 54 US-09-471-277-1906
25 18 3.4 224 18 US-08-727-609-1623
26 18 3.4 232 23 US-08-853-285-433
27 18 3.4 232 57 US-60-017-601-433
28 18 3.4 238 39 US-09-100-454-358
29 18 3.4 238 66 US-60-051-749-358
30 18 3.4 244 25 US-08-900-077-1591
31 18 3.4 249 2 US-07-837-195C-1434
32 18 3.4 254 24 US-08-885-220B-3527
33 18 3.4 254 58 US-60-020-591-3527
34 18 3.4 266 70 US-60-084-055-3488
35 18 3.4 269 43 US-09-231-945-2711
36 18 3.4 272 41 US-09-169-546-655
37 18 3.4 275 70 US-60-084-055-3550
38 18 3.4 276 44 US-09-262-979-1350
39 18 3.4 280 72 US-60-090-928-1716
40 18 3.4 284 69 US-60-082-567-1009
41 18 3.4 284 37 US-09-065-511-1648
42 18 3.4 296 64 US-60-044-939-1648
43 18 3.4 299 37 US-09-070-695A-1295
44 18 3.4 319 50 US-09-359-067-9002
45

```

# ALIGNMENTS

```

RESULT 1
US-09-215-435-76
; Sequence 76, Application US/09215435A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric
; APPLICANT: Bougueleret Lydie
; TITLE OF INVENTION: Extended cDNAs
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215,435A
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/069,957
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 76
; LENGTH: 526
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 22..318
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 22..93
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 4.6
; OTHER INFORMATION: seq FFIFCSLNTLLG/GV
; FEATURE:
; NAME/KEY: polyA_signal

```

```

Sequence 583, App
Sequence 6596, App
Sequence 829, App
Sequence 9978, App
Sequence 8101, App
Sequence 2025, App
Sequence 458, App
Sequence 14446, A
Sequence 978, App
Sequence 2235, App
Sequence 1906, App
Sequence 1623, App
Sequence 433, App
Sequence 358, App
Sequence 358, App
Sequence 1591, App
Sequence 1434, App
Sequence 3527, App
Sequence 3527, App
Sequence 3488, App
Sequence 2711, App
Sequence 655, App
Sequence 3550, App
Sequence 1350, App
Sequence 4442, App
Sequence 1716, App
Sequence 1009, App
Sequence 1648, App
Sequence 1648, App
Sequence 1295, App
Sequence 9002, App

```

```

; LOCATION: 497..502
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 516..526
US-09-215-435-76

Query Match
Best Local Similarity 100.0%; Score 526; DB 42; Length 526;
Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgcctgctgtgtgcacatgaagtgcacagctgggagattcttcttaagattcttc 60
Db 1 ctgcctgctgtgtgcacatgaagtgcacagctgggagattcttcttaagattcttc 60
QY 61 atctctgtctattgaataccctgttattgggtgttaataaaaattgcggagaagata 120
Db 61 atctctgtctattgaataccctgttattgggtgttaataaaaattgcggagaagata 120
QY 121 tctggagacctcaaaagatccctgcacatggacatgaattttggagctgctatgaatt 180
Db 121 tctggagacctcaaaagatccctgcacatggacatgaattttggagctgctatgaatt 180
QY 181 cactttagattttctacaacagacacctccaaaagatgtgaacctttgtcttcctcgc 240
Db 181 cactttagattttctacaacagacacctccaaaagatgtgaacctttgtcttcctcgc 240
QY 241 tgaatggcaaccttaacaaactcaagcttaaaatagaaactggaagtagcctgtgtgca 300
Db 241 tgaatggcaaccttaacaaactcaagcttaaaatagaaactggaagtagcctgtgtgca 300
QY 301 aaatacaaacaccagggtgagaggatgtgaactcatgaagtgtctgtcgcaccatccg 360
Db 301 aaatacaaacaccagggtgagaggatgtgaactcatgaagtgtctgtcgcaccatccg 360
QY 361 aaataaagacacaaagaaattcagactgatttgaataatttgaataatttccataatgc 420
Db 361 aaataaagacacaaagaaattcagactgatttgaataatttgaataatttccataatgc 420
QY 421 ttttaagcttccatagtgttgcattttctgcaccttagtttgccttcttctggaataaa 480
Db 421 ttttaagcttccatagtgttgcattttctgcaccttagtttgccttcttctggaataaa 480
QY 481 ctgtatgattcattagaatgaagagttcttctgtcctcaaaaaaa 526
Db 481 ctgtatgattcattagaatgaagagttcttctgtcctcaaaaaaa 526

```

## RESULT 2

```

US-09-215-435-303
; Sequence 303, Application US/09215435A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric
; APPLICANT: Bougueleret Lydie
; TITLE OF INVENTION: Extended cDNAs
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215,435A
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/069,957
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 303
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```
FEATURE:
NAME/KEY: CDS
LOCATION: 32..328
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 32..103
OTHER INFORMATION: Von Heijne matrix
OTHER INFORMATION: score 4.59999990463257
OTHER INFORMATION: seq FFIFCSLNTLLG/GV
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 508..513
FEATURE:
NAME/KEY: polyA_site
LOCATION: 528..539
US-09-215-435-303

Query Match      53.4%; Score 281; DB 42; Length 539;
Best Local Similarity 99.5%; Pred. No. 2.1e-131;
Matches 381; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctgctgctgctgctgaccatgaagctgccaaagctgggatttcttctaagattcttc 60
Db 11 ctgctgctgctgctgaccatgaagctgccaaagctgggatttcttctaagattcttc 70

Qy 61 atctctgctcatgaatacaccctgtattgggtggtgttaataaaattcgcgagaagata 120
Db 71 atctctgctcatgaatacaccctgtattgggtggtgttaataaaattcgcgagaagata 130

Qy 121 tgtggagaccctcaaaagatccctgcaaatggacatgaatttggagctgctatgaagtt 180
Db 131 tgtggagaccctcaaaagatccctgcaaatggacatgaatttggagctgctatgaagtt 190

Qy 181 cacttagatatttctacaacagaccctccaaagatgtgaaactttgtcttcccgcc 240
Db 191 cacttagatatttctacaacagaccctccaaagatgtgaaactttgtcttcccgcc 250

Qy 241 tgaatggcaaccttaacaacttcaagcttaaaatagaacgtgaagctgctgtgttgc 300
Db 251 tgaatggcaaccttaacaacttcaagcttaaaatagaacgtgaagctgctgtgttgc 310

Qy 301 aaatacaaacaccagagtgagagatgtaactaatgaattgctgctgcaccatccg 360
Db 311 aaatacaaacaccagagtgagagatgtaactaatgaattgctgctgcaccatccg 370

Qy 361 aaataaagacacaaagaaattca 383
Db 371 aaataaagacacaaagaaattca 393

RESULT 3
US-60-069-957-136
; Sequence 136, Application US/60069957
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste Dumas Milne
; APPLICANT: Duclert, Aymeric
; TITLE OF INVENTION: EXTENDED CDNAS FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 381
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/069,957
```

```
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.019PR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 136:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 539 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Testis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 32..328
; NAME/KEY: sig_peptide
; LOCATION: 32..103
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 4.6
; OTHER INFORMATION: seq FFIFCSLNTLLG/GV
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 508..513
; NAME/KEY: polyA
; LOCATION: 528..539
; IDENTIFICATION METHOD: blastn
US-60-069-957-136
```

```
Query Match      53.4%; Score 281; DB 67; Length 539;
Best Local Similarity 99.5%; Pred. No. 2.1e-131;
Matches 381; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctgctgctgctgctgaccatgaagctgccaaagctgggatttcttctaagattcttc 60
Db 11 ctgctgctgctgctgaccatgaagctgccaaagctgggatttcttctaagattcttc 70

Qy 61 atctctgctcatgaatacaccctgtattgggtggtgttaataaaattcgcgagaagata 120
Db 71 atctctgctcatgaatacaccctgtattgggtggtgttaataaaattcgcgagaagata 130

Qy 121 tgtggagaccctcaaaagatccctgcaaatggacatgaatttggagctgctatgaagtt 180
Db 131 tgtggagaccctcaaaagatccctgcaaatggacatgaatttggagctgctatgaagtt 190

Qy 181 cacttagatatttctacaacagaccctccaaagatgtgaaactttgtcttcccgcc 240
Db 191 cacttagatatttctacaacagaccctccaaagatgtgaaactttgtcttcccgcc 250

Qy 241 tgaatggcaaccttaacaacttcaagcttaaaatagaacgtgaagctgctgtgttgc 300
Db 251 tgaatggcaaccttaacaacttcaagcttaaaatagaacgtgaagctgctgtgttgc 310

Qy 301 aaatacaaacaccagagtgagagatgtaactaatgaattgctgctgcaccatccg 360
Db 311 aaatacaaacaccagagtgagagatgtaactaatgaattgctgctgcaccatccg 370

Qy 361 aaataaagacacaaagaaattca 383
Db 371 aaataaagacacaaagaaattca 393

RESULT 4
US-08-905-279-148
```

```
; Sequence 148, Application US/08905279
; GENERAL INFORMATION:
; APPLICANT: Genset SA
; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
; TITLE OF INVENTION: EXPRESSED IN TESTIS AND OTHER TISSUES
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,279
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.011A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 271 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Testis
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 32..103
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 4.6
; OTHER INFORMATION: seq FIFCSLNTLLIG/GV
US-08-905-279-148

Query Match 45.1%; Score 237; DB 25; Length 271;
Best Local Similarity 100.0%; Pred. No. 3.4e-109;
Matches 237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgcctgctcttgcgcacatgaagctgcgaagctggcgaagctggatttcttcttaagattcttc 60
Db 11 CTGCCTGCTCTTGCCTGCACCATGAAGCTGCACCAAGCTGGGATTTCTTCAAGATTCCTC 70

QY 61 attcttgcctcatgaataccctgttatggcggttaataaaatttcgcgaagata 120
Db 71 ATCTTCTGCTCATGATATACCCCTGTTATTCGGTGGTGTAAATAAATTCGGGAGAAGATA 130

QY 121 tctggagacctcaagaatccctgcgaatggacatgaattttggaagctgctatgaagtt 180
Db 131 TGTGGAGACCTCAAGAATCCCTGCAATTTGGACATGAATTTTGAAGCTGCTATGAAGTT 190

QY 181 cacttagatattctacacagaaacctccaaaagatggaacttttctctctcc 237
Db 191 CACTTTAGATATTCTACACAGAACTCCCAAAAGATGTGAACACTTTTGTCTTCTCC 247

RESULT 5
US-09-248-796-9251/c
; Sequence 9251, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 9251
; LENGTH: 762
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796-9251

Query Match 4.0%; Score 21; DB 43; Length 1101;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 389 atttgaatactttgtaatat 409
Db 483 ATTTGAAATCTTTGTAATAT 463

RESULT 6
US-60-096-409-9251/c
; Sequence 9251, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096,409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 9251
; LENGTH: 762
; TYPE: DNA
; ORGANISM: Candida albicans
US-60-096-409-9251

Query Match 4.0%; Score 21; DB 73; Length 762;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 389 atttgaatactttgtaatat 409
Db 483 ATTTGAAATCTTTGTAATAT 463

RESULT 7
US-09-248-796-2842/c
; Sequence 2842, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 2842
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796-2842

Query Match 4.0%; Score 21; DB 43; Length 1101;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 389 atttgaatactttgtaatat 409
```

Db 351 ATTGGAATCTTTGTAATAT 331  
|||||  
RESULT 8  
US-60-096-409-2842/c  
; Sequence 2842, Application US/60096409A  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; FILE REFERENCE: GTC-016P  
; CURRENT APPLICATION NUMBER: US/60/096,409A  
; CURRENT FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28206  
; SEQ ID NO 2842  
; LENGTH: 1101  
; TYPE: DNA  
; ORGANISM: Candida albicans  
US-60-096-409-2842

Query Match 4.0%; Score 21; DB 73; Length 1101;  
Best Local Similarity 100.0%; Pred. No. 5.3; Mismatches 0; Indels 0; Gaps 0;  
Matches 21; Conservative 0;  
QY 389 atttgaatctttgtaatat 409  
|||||  
Db 351 ATTGGAATCTTTGTAATAT 331  
RESULT 9  
US-60-145-485-7430/c  
; Sequence 7430, Application US/60145485  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; APPLICANT: Deng, Mollan  
; APPLICANT: Fisher, Dane K.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; FILE REFERENCE: 38-21(51229)A  
; CURRENT APPLICATION NUMBER: US/60/145,485  
; CURRENT FILING DATE: 1999-07-23  
; NUMBER OF SEQ ID NOS: 9366  
; SEQ ID NO 7430  
; LENGTH: 378  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: unsure at all n locations  
; OTHER INFORMATION: Clone ID: LIB3150-092-P2-K1-D4  
US-60-145-485-7430

Query Match 3.8%; Score 20; DB 84; Length 378;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 507 cttctgtcaaaaaaaa 526  
|||||  
Db 49 CTTCTGTCAAAAAAAA 30  
RESULT 10  
US-60-138-437-754  
; Sequence 754, Application US/60128437  
; GENERAL INFORMATION:  
; APPLICANT: Conner, Timothy W.  
; APPLICANT: Masucci, James D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(15725)A

; CURRENT APPLICATION NUMBER: US/60/128,437  
; CURRENT FILING DATE: 1999-04-06  
; NUMBER OF SEQ ID NOS: 5533  
; SEQ ID NO 754  
; LENGTH: 383  
; TYPE: DNA  
; ORGANISM: Triticum aestivum  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB103-007-Q1-E1-E6  
US-60-128-437-754

Query Match 3.8%; Score 20; DB 79; Length 383;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 507 cttctgtcaaaaaaaa 526  
|||||  
Db 215 cttctgtcaaaaaaaa 234

RESULT 11  
US-60-168-611-752/c  
; Sequence 752, Application US/60168611  
; GENERAL INFORMATION:  
; APPLICANT: Hodgson, David M. Jones, Anissa L.  
; APPLICANT: Lincoln, Stephen E. Yu, Jimmy Y.  
; APPLICANT: Russo, Frank D. Panzer, Scott R.  
; APPLICANT: Spiro, Peter A. Roseberry, Ann M.  
; APPLICANT: Banville, Steve C. Wright, Rachel J.  
; APPLICANT: Bratcher, Shawn R. Chen, Wensheng  
; APPLICANT: Dufour, Gerard J. Liu, Tommy  
; APPLICANT: Cohen, Howard J. Yap, Pierre E.  
; APPLICANT: Shah, Bruce Amshey, Stefan  
; APPLICANT: Chalup, Michael S. Fong, Willy Tuen  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: SECRETORY MOLECULES  
; FILE REFERENCE: PT-0111 P  
; CURRENT APPLICATION NUMBER: US/60/168,611  
; CURRENT FILING DATE: 1999-12-02  
; NUMBER OF SEQ ID NOS: 4000  
; SOFTWARE: PERL Program  
; SEQ ID NO 752  
; LENGTH: 499  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No: 985509.1  
US-60-168-611-752

Query Match 3.8%; Score 20; DB 86; Length 499;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 212 aaagatgtgaaactttgtc 231  
|||||  
Db 455 AAAGATGTGAACCTTTGTC 436  
RESULT 12  
US-09-041-720-3385/c  
; Sequence 3385, Application US/09041720  
; GENERAL INFORMATION:  
; APPLICANT: Gooding, Douglas H.  
; APPLICANT: Stuve, Laura L.  
; APPLICANT: Stuart, Susan G.  
; APPLICANT: Ito, Laura Y.  
; APPLICANT: Akerblom, Ingrid E.  
; APPLICANT: Deleageane, Angelo M.  
; APPLICANT: Naughton, Rebecca E.

APPLICANT: Klingler, Tod M.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM  
NUMBER OF SEQUENCES: 4388  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/041,720  
FILING DATE: HEREWITH  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/040,197  
FILING DATE: MARCH 14, 1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PD-0333 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 3385:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 275 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
IMMEDIATE SOURCE:  
CLONE: 3974221H1  
US-09-041-720-3385

Query Match 3.6%; Score 19; DB 36; Length 275;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 tctacacagaaacctccaa 212  
|||||  
DB 19 TCTACACAGAACCTCCAA 1

RESULT 13  
US-08-822-285-583  
Sequence 583, Application US/08822285  
GENERAL INFORMATION:  
APPLICANT: Gooding, Douglas H.  
APPLICANT: Stuve, Laura L.  
APPLICANT: Stuart, Susan G.  
APPLICANT: Ito, Laura Y.  
APPLICANT: Akerblom, Ingrid E.  
APPLICANT: Delegeane, Angelo M.  
APPLICANT: Naughton, Rebecca E.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM  
NUMBER OF SEQUENCES: 4210  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA

ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/822,285  
FILING DATE: Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/013,696  
FILING DATE: March 20, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PD-0133 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 855-0555  
TELEFAX: (415) 845-4166  
INFORMATION FOR SEQ ID NO: 583:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
IMMEDIATE SOURCE:  
CLONE: 894698  
US-08-822-285-583

Query Match 3.6%; Score 19; DB 22; Length 303;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 aaaccaccgaggtgagag 325  
|||||  
DB 278 AAACCACCGAGGTGAGG 296

RESULT 14  
US-60-013-696-583  
Sequence 583, Application US/60013696  
GENERAL INFORMATION:  
APPLICANT: Gooding, Douglas H.  
APPLICANT: Stuve, Laura L.  
APPLICANT: Stuart, Susan G.  
APPLICANT: Ito, Laura Y.  
APPLICANT: Akerblom, Ingrid E.  
APPLICANT: Delegeane, Angelo M.  
APPLICANT: Naughton, Rebecca E.  
APPLICANT: Klingler, Tod M.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM  
NUMBER OF SEQUENCES: 6327  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/60/013,696  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

```

; NAME: LUTHER, BARBARA J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PD-0133P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 855-0555
; TELEFAX: (415) 852-0195
; INFORMATION FOR SEQ ID NO: 583:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: 894698
; US-60-013-696-583

```

```

Query Match      3.6%; Score 19; DB 56; Length 303;
Best Local Similarity 100.0%; Pred.No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 307 aaaccaccgaggtgagagg 325
    |||||
Db 278 AAACCACCGAGGTGAGAGG 296

```

```

RESULT 15
US-60-145-485-6596/c
; Sequence 6596, Application US/60145485
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordin
; APPLICANT: Deng, Mollan
; APPLICANT: Fisher, Dane K.
; APPLICANT: Miller, Philip W.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 38-21(51229)A
; CURRENT APPLICATION NUMBER: US/60/145.485
; CURRENT FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 9366
; SEQ ID NO 6596
; LENGTH: 312
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3150-082-P2-N2-B6
US-60-145-485-6596

```

```

Query Match      3.6%; Score 19; DB 84; Length 312;
Best Local Similarity 100.0%; Pred.No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 508 ttctgtcaaaaaaaaaa 526
    |||||
Db 33 TTCTGTCAAAAAAAAAA 15

```

Search completed: May 1, 2000, 19:59:38  
Job time: 21962 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 14:44:17 ; Search time 3022.95 Seconds  
(without alignments)  
656.974 Million cell updates/sec

Title: us-09-215-435-76  
Perfect score: 526  
Sequence: 1 ctgctgtctgtctgtgcac.....cttctgtcaaaaaaaaaa 526

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4538634 seqs, 1887831982 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

- EST:\*
- 1: em\_est1:\*
  - 2: em\_est2:\*
  - 3: em\_est3:\*
  - 4: em\_est4:\*
  - 5: em\_est5:\*
  - 6: em\_est6:\*
  - 7: em\_est7:\*
  - 8: em\_est8:\*
  - 9: em\_est9:\*
  - 10: em\_est10:\*
  - 11: em\_est11:\*
  - 12: em\_est12:\*
  - 13: em\_est13:\*
  - 14: em\_est14:\*
  - 15: em\_est15:\*
  - 16: em\_est16:\*
  - 17: em\_est17:\*
  - 18: em\_est18:\*
  - 19: em\_est19:\*
  - 20: gb\_est1:\*
  - 21: gb\_est2:\*
  - 22: gb\_est3:\*
  - 23: gb\_est4:\*
  - 24: gb\_est5:\*
  - 25: gb\_est6:\*
  - 26: gb\_est7:\*
  - 27: gb\_est8:\*
  - 28: gb\_est9:\*
  - 29: gb\_est10:\*
  - 30: gb\_est11:\*
  - 31: gb\_est12:\*
  - 32: gb\_est13:\*
  - 33: gb\_est14:\*
  - 34: gb\_est15:\*
  - 35: gb\_est16:\*
  - 36: gb\_est17:\*
  - 37: gb\_est18:\*
  - 38: gb\_est19:\*
  - 39: gb\_est20:\*
  - 40: gb\_est21:\*
  - 41: gb\_est22:\*
  - 42: gb\_est23:\*
  - 43: gb\_est24:\*

- 44: gb\_est25:\*
- 45: gb\_est26:\*
- 46: gb\_est27:\*
- 47: gb\_est28:\*
- 48: gb\_est29:\*
- 49: gb\_est30:\*
- 50: gb\_est31:\*
- 51: gb\_est32:\*
- 52: em\_est20:\*
- 53: em\_est21:\*
- 54: em\_est22:\*
- 55: em\_est23:\*
- 56: em\_est24:\*
- 57: em\_est25:\*
- 58: em\_est26:\*
- 59: gb\_est33:\*
- 60: gb\_est34:\*
- 61: gb\_est35:\*
- 62: gb\_est36:\*
- 63: gb\_est37:\*
- 64: gb\_est38:\*
- 65: em\_est27:\*
- 66: em\_est28:\*
- 67: em\_est29:\*
- 68: em\_est30:\*
- 69: gb\_est39:\*
- 70: gb\_est40:\*
- 71: gb\_est41:\*
- 72: gb\_est42:\*
- 73: gb\_est43:\*
- 74: gb\_est44:\*
- 75: em\_est31:\*
- 76: em\_est32:\*
- 77: em\_est33:\*
- 78: em\_est34:\*
- 79: gb\_gss1:\*
- 80: gb\_gss2:\*
- 81: gb\_gss3:\*
- 82: gb\_gss4:\*
- 83: em\_gss1:\*
- 84: em\_gss2:\*
- 85: em\_gss3:\*
- 86: em\_gss4:\*
- 87: gb\_gss5:\*
- 88: gb\_gss6:\*
- 89: gb\_gss7:\*
- 90: gb\_gss8:\*
- 91: gb\_gss9:\*
- 92: em\_gss5:\*
- 93: em\_gss6:\*
- 94: em\_gss7:\*
- 95: em\_gss8:\*
- 96: em\_gss9:\*
- 97: em\_gss10:\*
- 98: em\_gss11:\*
- 99: gb\_gss10:\*
- 100: gb\_gss11:\*
- 101: em\_gss12:\*
- 102: gb\_gss12:\*
- 103: gb\_gss13:\*
- 104: gb\_gss14:\*
- 105: gb\_gss15:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result				Query		Match		Length		DB		ID		Description	
No.	Score	Match	Length	DB	ID	No.	Score	Match	Length	DB	ID	No.	Score	Match	Description
c	1	95	18.1	409	81	B46318	HS-1063-B1-								

21 4.0 282 70 AW156800 se31a04.y  
21 4.0 354 51 AT1737290 606038E05  
21 4.0 739 100 AQ326122 nbxb00025E  
5 20 3.8 196 64 AW086346 xc72f09.x  
6 20 3.8 396 79 CNS00PVB  
7 20 3.8 421 51 AT1712733 Arabidops  
8 20 3.8 437 34 AA535639 U1-R-AF1-  
9 19 3.6 100 51 AT1736037 U1-R-AF1-  
10 19 3.6 220 34 AA516245 ng66c04.s  
11 19 3.6 279 59 AV089957 AV089957  
12 19 3.6 287 69 AW120603 U1-N-BH2.  
13 19 3.6 306 49 AU057492 AU057492  
14 19 3.6 310 27 AA050528  
15 19 3.6 337 43 C99144 C99144  
16 19 3.6 355 41 AU044930 U1-R-C1-k  
17 19 3.6 377 29 AA170688 ms85f10.i  
18 19 3.6 377 44 AU1322904 ms85f10.y  
19 19 3.6 393 49 AU057645 AU057645  
20 19 3.6 396 64 AW047018 U1-N-BH1-  
21 19 3.6 406 64 AW076804 fJ04a07.x  
22 19 3.6 407 100 AQ268428 RPCi11-74  
23 19 3.6 409 45 AI394946 MA001784.  
24 19 3.6 412 42 AI093417 qb03h06.x  
25 19 3.6 418 48 AI575376 U1-R-G0-u  
26 19 3.6 421 64 AW022845 df45a04.y  
27 19 3.6 438 51 AT1739601 U1-R-AF1-  
28 19 3.6 429 51 AT1714089 U1-R-AF1-  
29 19 3.6 431 46 AI431276 ar57b10.x  
30 19 3.6 443 50 AT1705147 U1-R-G0-u  
31 19 3.6 448 69 AW121827 U1-N-BH2.  
32 19 3.6 475 37 AA682389 zJ86c03.s  
33 19 3.6 475 60 AT1794362 sb74b12.y  
34 19 3.6 476 40 AA982860 ub59f11.i  
35 19 3.6 478 42 AI136568 U1-R-C2p-  
36 19 3.6 479 64 AW017059 EST72404  
37 19 3.6 482 40 AA960570 ub59f11.s  
38 19 3.6 509 45 AI395770 MA005120.  
39 19 3.6 515 47 AI511115 U1-R-BF0-  
40 19 3.6 562 61 AI819349 wg61a10.x  
41 19 3.6 565 48 AI615190 486094C11  
42 19 3.6 568 50 AI678669 tu58e02.x  
43 19 3.6 593 51 AT1737778 606041806  
44 19 3.6 599 51 AT1738157 606046D11  
45 19 3.6 649 49 AI649954 486099601

## ALIGNMENTS

RESULT 1  
B46318/c  
LOCUS  
DEFINITION HS-1063-B1-A10-MR.abi CIT Human Genomic Sperm Library C Homo  
sapientis genomic clone Plate-CT 796 Col=19 Row=B, genomic survey  
sequence.  
B46318  
B46318.1 GI:2551152  
GSS.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 409)  
Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,  
Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.  
Construction of a Characterized Clone Resource for Genomic  
Sequencing: Generation and Preliminary Analysis of 20,000 Sequence  
Tagged Connectors  
Unpublished (1997)  
Contact: Mahairas GG, Zackrone KD, Hood L  
University of Washington  
Seattle, WA 98195, USA  
Tel: (206) 616-8744

21 4.0 282 70 AW156800 se31a04.y  
21 4.0 354 51 AT1737290 606038E05  
21 4.0 739 100 AQ326122 nbxb00025E  
5 20 3.8 196 64 AW086346 xc72f09.x  
6 20 3.8 396 79 CNS00PVB  
7 20 3.8 421 51 AT1712733 Arabidops  
8 20 3.8 437 34 AA535639 U1-R-AF1-  
9 19 3.6 100 51 AT1736037 U1-R-AF1-  
10 19 3.6 220 34 AA516245 ng66c04.s  
11 19 3.6 279 59 AV089957 AV089957  
12 19 3.6 287 69 AW120603 U1-N-BH2.  
13 19 3.6 306 49 AU057492 AU057492  
14 19 3.6 310 27 AA050528  
15 19 3.6 337 43 C99144 C99144  
16 19 3.6 355 41 AU044930 U1-R-C1-k  
17 19 3.6 377 29 AA170688 ms85f10.i  
18 19 3.6 377 44 AU1322904 ms85f10.y  
19 19 3.6 393 49 AU057645 AU057645  
20 19 3.6 396 64 AW047018 U1-N-BH1-  
21 19 3.6 406 64 AW076804 fJ04a07.x  
22 19 3.6 407 100 AQ268428 RPCi11-74  
23 19 3.6 409 45 AI394946 MA001784.  
24 19 3.6 412 42 AI093417 qb03h06.x  
25 19 3.6 418 48 AI575376 U1-R-G0-u  
26 19 3.6 421 64 AW022845 df45a04.y  
27 19 3.6 438 51 AT1739601 U1-R-AF1-  
28 19 3.6 429 51 AT1714089 U1-R-AF1-  
29 19 3.6 431 46 AI431276 ar57b10.x  
30 19 3.6 443 50 AT1705147 U1-R-G0-u  
31 19 3.6 448 69 AW121827 U1-N-BH2.  
32 19 3.6 475 37 AA682389 zJ86c03.s  
33 19 3.6 475 60 AT1794362 sb74b12.y  
34 19 3.6 476 40 AA982860 ub59f11.i  
35 19 3.6 478 42 AI136568 U1-R-C2p-  
36 19 3.6 479 64 AW017059 EST72404  
37 19 3.6 482 40 AA960570 ub59f11.s  
38 19 3.6 509 45 AI395770 MA005120.  
39 19 3.6 515 47 AI511115 U1-R-BF0-  
40 19 3.6 562 61 AI819349 wg61a10.x  
41 19 3.6 565 48 AI615190 486094C11  
42 19 3.6 568 50 AI678669 tu58e02.x  
43 19 3.6 593 51 AT1737778 606041806  
44 19 3.6 599 51 AT1738157 606046D11  
45 19 3.6 649 49 AI649954 486099601

FEATURES  
source  
1. 409  
Location/Qualifiers  
High quality sequence stop: 409.  
Class: BAC ends  
Plate: CT 796 row: B column: 19  
Sequence Tagged Connector  
Email: kzackrone@u.washington.edu  
Fax: (206) 685-7301

BASE COUNT 126 a 86 c 53 g 133 t 1 others  
ORIGIN  
E-Coli DH10B"  
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in  
/db\_xref="taxon:9606"  
/clone\_lib="CIT Human Genomic Sperm Library C"  
/sex="M"  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

Query Match 18.1%; Score 95; DB 81; Length 409;  
Best Local Similarity 100.0%; Pred. No. 6.3e-37;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 315 gaggtgagaggatggaactcattgaagtgtctgtcaccatccgaataagacacaa 374  
|||||  
Db 135 GAGGTGAGAGGATGGAACATCATGAAGTTGTCTGTCACCATCGGAATAAGACACAA 76  
|||||  
QY 375 gaaattcagactgattggaattcattgtaatat 409  
|||||  
Db 75 GAAATTCAGACTGATTGGAATCTTTGTAATAT 41  
|||||

RESULT 2  
AW156800  
LOCUS  
DEFINITION se31a04.y1 Gm-cl015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
Gm-cl015-2791 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
Glycine.

1 (bases 1 to 282)  
Shoemaker,R., Keim,P., Vodkin,L., Erpeliding,J., Coryell,V.,  
Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,  
Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,  
Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,  
Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,  
McCann,R., Waterston,R. and Wilson,R.  
Public Soybean EST Project  
Unpublished (1999)  
On Dec 20, 1995 this sequence version replaced gi:1134791.  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available through: Genome Systems, Inc. 4633 World  
Parkway Circle St. Louis, Missouri 63134 For further information  
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)  
427-3324 or contact: clones@genomesystems.com or  
info@genomesystems.com web site: www.genomesystems.com  
Seq primer: -40RP from Gibco.  
Location/Qualifiers  
1. 282  
/db\_xref="taxon:3847"

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Unpublished (1999)  
On Dec 20, 1995 this sequence version replaced gi:1134791.  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available through: Genome Systems, Inc. 4633 World  
Parkway Circle St. Louis, Missouri 63134 For further information  
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)  
427-3324 or contact: clones@genomesystems.com or  
info@genomesystems.com web site: www.genomesystems.com  
Seq primer: -40RP from Gibco.  
Location/Qualifiers  
1. 282  
/db\_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl015-2791"  
 /clone\_lib="Gm-cl015"  
 /tissue\_type="Mature flowers, field grown plants"  
 /lab\_host="XL10-Gold"  
 /note="vector: pBluescript II XR; Site\_1: EcoRI; Site\_2: XhoI; This cDNA library was constructed from mRNA isolated from mature flowers of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelding."  
 BASE COUNT 159 a 27 c 28 g 68 t  
 ORIGIN

Query Match 4.0%; Score 21; DB 70; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 4.6;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 506 tctttctgtcaaaaaaaa 526  
 |||||  
 Db 172 TCTTCTGTCAAAAAAAA 192

RESULT 3  
 LOCUS A1737290/c 554 bp mRNA EST 14-JUN-1999  
 DEFINITION 606038E05.x1 606 - Ear tissue cDNA library from Schmidt lab Zea  
 mays cDNA, mRNA sequence.  
 ACCESSION A1737290  
 VERSION A1737290.1 GI:5058814  
 KEYWORDS EST.  
 SOURCE Zea mays.  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
 Poaceae; Zea.  
 REFERENCE 1 (bases 1 to 554)  
 AUTHORS Walbot,V.  
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University  
 JOURNAL Unpublished (1999)  
 COMMENT On May 18, 1998 this sequence version replaced gi:3138716.  
 Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu  
 Plate: 606038 row: E column: 05.  
 FEATURES Location/Qualifiers  
 1..554  
 /organism="Zea mays"  
 /cultivar="Ohio43"  
 /db\_xref="taxon:4577"  
 /clone\_lib="606 - Ear tissue cDNA library from Schmidt lab"  
 /tissue\_type="mixed"  
 /dev\_stage="before pollen shed"  
 /lab\_host="XL10L (Stratagene)"  
 /note="Organ: immature ear; Vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI; Mixed ear tissue cDNA library from Schmidt lab"

BASE COUNT 112 a 208 c 152 g 82 t  
 ORIGIN

Query Match 4.0%; Score 21; DB 51; Length 554;  
 Best Local Similarity 100.0%; Pred. No. 3.9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 506 tctttctgtcaaaaaaaa 526  
 |||||  
 Db 29 TCTTCTGTCAAAAAAAA 9

RESULT 4  
 LOCUS AQ326122 739 bp DNA GSS 08-JAN-1999  
 DEFINITION nbxb0025808f CUGI Rice BAC Library Oryza sativa genomic clone  
 nbxb0025808f, genomic survey sequence.  
 ACCESSION AQ326122  
 VERSION AQ326122.1 GI:4117972  
 KEYWORDS GSS.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
 Poaceae; Oryza.  
 REFERENCE 1 (bases 1 to 739)  
 AUTHORS Wing,R.A. and Dean,R.A.  
 TITLE A BAC End Sequencing Framework to Sequence the Rice Genome  
 JOURNAL Unpublished (1998)  
 COMMENT Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Seq primer: TAATACGACTCACTATAGGG  
 Class: BAC ends  
 High quality sequence stop: 437.

FEATURES source  
 1..739  
 Location/Qualifiers  
 /organism="Oryza sativa"  
 /strain="Japonica"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone="nbxb0025808f"  
 /clone\_lib="CUGI Rice BAC Library"  
 /tissue\_type="Leaf"  
 /lab\_host="E. coli DH10B"  
 /note="vector: pBel08AC11; Site\_1: HindIII; Site\_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 16,432 clones (doubly spotted), represent the whole library for colony screening."  
 BASE COUNT 233 a 156 c 146 g 204 t  
 ORIGIN

Query Match 4.0%; Score 21; DB 100; Length 739;  
 Best Local Similarity 100.0%; Pred. No. 3.6;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 144 caaattgacatgaattttg 164  
 Db 694 CAAATTGACATGAATTG 714

RESULT 5  
 AW086346 196 bp mRNA EST 14-OCT-1999  
 LOCUS xc72f09.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
 DEFINITION IMAGE:2589833 3', mRNA sequence.

ACCESSION AW086346  
 VERSION AW086346.1 GI:6041498  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 196)  
 AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Dec 20, 1995 this sequence version replaced gi:1134075.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov.  
 This clone is available royalty-free through LML; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40UP from Gibco.

FEATURES  
 Location/Qualifiers

1..196  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2589833"  
 /clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NBHL19W, testis NHT, and B-cell  
 NCI-CCAP GCBI) were mixed, and ss circles were made in  
 vitro. Following RAP purification, this DNA was used as  
 tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from pools of 5,000 clones made  
 from the same 3 libraries. The pools consisted of  
 I.M.A.G.E. clones 297480-302087, 682632-687239,  
 726408-728711, and 729096-731399. Subtraction by Bento  
 Soares and M. Fatima Bonaldo."

BASE COUNT 73 a 36 c 41 g 46 t

Query Match 3.8%; Score 20; DB 64; Length 196;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 507 ctttcgtcaaaaaa 526  
 Db 143 CTTTCGTCAAAAAA 162

RESULT 6  
 CNS00PVB/c 396 bp DNA GSS 28-JUN-1999  
 LOCUS Arabidopsis thaliana genome survey sequence SP6 end of BAC F9021 of  
 DEFINITION IGF library from strain Columbia of Arabidopsis thaliana, genomic  
 survey sequence.

ACCESSION AL085013  
 VERSION AL085013.1 GI:5286153  
 KEYWORDS GSS.  
 SOURCE thale cress.

ORGANISM

Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 Arabidopsis.

REFERENCE 1 (bases 1 to 396)

AUTHORS Salanoubat,M., Choise,N., Artiguenave,F., Brottier,P., Wincker,P.,  
 Samson,D., Saurin,W., Weissbach,J. and Quetier,F.

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 396)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr  
 - Web : www.genoscope.cns.fr)

FEATURES

Location/Qualifiers

1..396  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone\_lib="IGF"  
 /clone="F9021"  
 /note="end : SP6"

BASE COUNT 111 a 69 c 82 g 134 t

Query Match 3.8%; Score 20; DB 79; Length 396;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 387 tgatttgaaatcttggtaa 406

Db 267 TGATTTGAATCTTTGTA 248

RESULT 7

AW172733/c

LOCUS

DEFINITION

UI-R-AFI-auu-h-02-0-UI.s1 UI-R-AFI Rattus norvegicus cDNA clone

UI-R-AFI-auu-h-02-0-UI 3', mRNA sequence.

ACCESSION AW172733

VERSION AW172733.1

KEYWORDS GI:5016533

EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 421)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT On Jun 22, 1998 this sequence version replaced gi:3246985.

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dt track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dt track served to verify it as a clone from the

normalized AV canal at 15 dpc library cDNA Library Preparation:

M.B. Soares Lab Clone distribution (www.resgen.com)

through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA-Yes.

FEATURES Location/Qualifiers

```

source
1. 421
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-AF1-aau-h-02-0-UI"
/clone_lib="UI-R-AF1"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AF1
library is a normalized library constructed from 15 dpc
rat atrioventricular (AV) canal. The tag is a string of 5
nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6:
791-806, 1996. Tissue provided by Jim Lin, Department of
Biology, University of Iowa.
TAG_LIB=UI-R-AF1
TAG_TISSUE=AV canal at 15 dpc
TAG_SEQ=GAAGG"
BASE COUNT      132 a      86 c      103 g      100 t
ORIGIN

Query Match      3.8%; Score 20; DB 51; Length 421;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 cttctgtcaaaaaaaaaa 526
|||||
Db 26 CTTTCTGTCAAAAAAAA 7

RESULT 8
AA535639/c
LOCUS      AA535639      437 bp      mRNA      EST      21-AUG-1997
DEFINITION n175h03.s1 NCI_CGAP_p10 Homo sapiens cDNA clone IMAGE:998357
similar to gb:M64241 QM PROTEIN (HUMAN)); mRNA sequence.
ACCESSION  AA535639
VERSION     AA535639.1 GI:2279892
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 437)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On May 9, 1995 this sequence version replaced gi:802538.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquil,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1022 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 55.
Location/Qualifiers
1. 437
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:998357"
/clone_lib="NCI_CGAP_p10"
/sex="male"

FEATURES
source
1. 100
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl007-309"
/clone_lib="Gm-cl007"
/lab_host="DH10B"
/notes="Vector: pSPORT1; Site_1: SalI; Site_2: NotI. This
cDNA library was constructed from mRNA isolated from
immature cotyledons of greenhouse grown plants (individual
seed fresh weight of 100-300mg). The library was prepared
using the Life Technologies pSuperScript cDNA library
construction kit. Complementary DNA was synthesized from
mRNA using a poly (dT) sequence with a NotI restriction
/tissue_type="invasive prostate tumor"
/lab_host="DH10B"
/notes="Organ: prostate; Vector: pAMP10; mRNA made from
invasive prostate tumor, cDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Library made by D. Krizman,
NIH."
BASE COUNT      101 a      82 c      120 g      134 t
ORIGIN

Query Match      3.8%; Score 20; DB 34; Length 437;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 cttctgtcaaaaaaaaaa 526
|||||
Db 31 CTTTCTGTCAAAAAAAA 12

RESULT 9
A1736037
LOCUS      A1736037      100 bp      mRNA      EST      06-DEC-1999
DEFINITION sb22all.v1 Gm-cl007 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl007-309 5', mRNA sequence.
ACCESSION  A1736037
VERSION     A1736037.1 GI:5057561
KEYWORDS    EST.
SOURCE      soybean.
ORGANISM    Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
Glycine.
1 (bases 1 to 100)
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V.,
Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,
Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,
Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N.,
Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
McCann, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
On May 16, 1998 this sequence version replaced gi:3137461.
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Putative full length read
vector to vector length is 101
Seq primer: -40RP from Gibco.
Location/Qualifiers
1. 100
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl007-309"
/clone_lib="Gm-cl007"
/lab_host="DH10B"
/notes="Vector: pSPORT1; Site_1: SalI; Site_2: NotI. This
cDNA library was constructed from mRNA isolated from
immature cotyledons of greenhouse grown plants (individual
seed fresh weight of 100-300mg). The library was prepared
using the Life Technologies pSuperScript cDNA library
construction kit. Complementary DNA was synthesized from
mRNA using a poly (dT) sequence with a NotI restriction

```

site. Sali linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-Sali restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E. coli ElectroMax DH10B host cells. This library was constructed by Dr. Lila Vodkin and Dr. Anu Khanna."

BASE COUNT 48 a 15 c 15 g 21 t 1 others

Query Match 3.6%; Score 19; DB 51; Length 100;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 508 ttctgtcaaaaaa 526  
|||||  
Db 73 TTTCTGTCAAAAA 91

RESULT 10  
AA516245/c  
LOCUS  
DEFINITION ng66c04.s1 NCI\_CGAP\_Lip2 Homo sapiens cDNA clone IMAGE:939750, mRNA  
sequence.  
ACCESSION AA516245  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM human.

AA516245 220 bp mRNA EST 19-AUG-1997

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CCGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 705 Std Error: 0.00  
Seq primer: 40ml3 fwd. ET from Amersham  
High quality sequence stop: 209.

FEATURES  
source

Location/Qualifiers  
1..220  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:939750"  
/clone\_lib="NCI\_CGAP\_Lip2"  
/tissue\_type="liposarcoma"  
/lab\_host="DH10B"  
/note="Vector: pAMP10; mRNA made from liposarcoma, cDNA made by oligo-dT priming. Non-directionally cloned.  
Size selected on agarose gel, average insert size 600 bp.  
Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."

BASE COUNT 68 a 32 c 65 g 55 t

Query Match 3.6%; Score 19; DB 34; Length 220;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 194 tctacaacagaacctccaa 212  
|||||  
Db 58 TCTACACAGAACCTCCAA 40

RESULT 11  
AV089957/c

LOCUS  
DEFINITION AV089957 Mus musculus tongue C57BL/6J adult Mus musculus cDNA clone  
2310046120, mRNA sequence.

ACCESSION AV089957  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS

Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Alizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
RIKEN Mouse ESTs  
Unpublished (1999)  
On May 18, 1998 this sequence version replaced gi:3137799.

TITLE  
JOURNAL  
COMMENT

Contact: Chie Owa  
Genome Science Laboratory  
RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-9145  
Fax: 81-298-36-9098  
Email: genome-res@tc.riken.go.jp

Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (http://genome.etc.riken.go.jp) for further details.

FEATURES  
source

Location/Qualifiers  
1..279  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="2310046120"  
/clone\_lib="Mus musculus tongue C57BL/6J adult"  
/sex="male"  
/tissue\_type="tongue"  
/dev\_stage="adult"

BASE COUNT 86 a 38 c 37 g 118 t

Query Match 3.6%; Score 19; DB 59; Length 279;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 508 ttctgtcaaaaaa 526  
|||||  
Db 79 TTTCTGTCAAAAA 61

RESULT 12  
AW120603/c

LOCUS  
DEFINITION AW120603 287 bp mRNA EST  
UI-M-BH2.3-ann-e-06-0-UI.s1 NIH\_BMAP\_M\_S3.3 Mus musculus cDNA clone  
UI-M-BH2.3-ann-e-06-0-UI 3', mRNA sequence.

ACCESSION AW120603  
VERSION

AW120603.1 GI:6095936

KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
TITLE 1 (bases 1 to 287)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 9704477  
COMMENT On May 18, 1998 this sequence version replaced gi:3137347.  
Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: MEST@mail.nih.gov

The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to identify it as a clone from the  
normalized hypothalamus library cDNA library Preparation: M.B.  
Soares lab Clone distribution: NIH BMAP cDNA clones will be made  
available by the means that is soon to be determined. When NIH  
determines the means for distribution of the BMAP cDNA clones, this  
record will be updated accordingly when that means is determined.  
The following repetitive elements were found in this cDNA sequence:  
1-51, >AT\_rich#Low\_complexity  
Seq primer: M13 Forward  
POLYA=Yes.

FEATURES  
source

Location/Qualifiers  
1..287  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="NIH\_BMAP\_M\_S3.3"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pTT3D-pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
NIH\_BMAP\_M\_S3.3 library is a subtracted library of a  
series, ultimately derived from a mixture of individually  
tagged normalized libraries from ten regions of the mouse  
brain (cerebellum, brain stems, olfactory bulbs, pineal  
hypothalamus, cortex, amygdala, basal ganglia, pineal  
gland, striatum, hippocampus) after a series of  
subtractions to reduce the representation of cDNAs from  
which ESTs had already been generated. The following  
serially subtracted libraries were generated in this  
process: NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1.  
The subtracted library (NIH\_BMAP\_M\_S3.3) was constructed  
as follows: PCR amplified cDNA inserts from NIH\_BMAP\_M\_S2  
clones from which 3' ESTs had been derived was used as a  
driver in a hybridization with the NIH\_BMAP\_M\_S2 library  
in the form of single-stranded circles. The remaining  
single-stranded circles (subtracted library) was purified  
by hydroxyapatite column chromatography, converted to  
double-stranded circles and electroporated into DH10B  
bacteria (Life Technologies) to generate the  
NIH\_BMAP\_M\_S3.3 library. This procedure has been  
previously described (Bonaldo, Lennon and Soares, Genome  
Research 6:791-806, 1996)  
TAG\_LIB=NIH\_BMAP\_M\_S3.3  
TAG\_TISSUE=hypothalamus  
TAG\_SEQ=CGGAA"

BASE COUNT  
ORIGIN

82 a 42 c 46 g 117 t

Query Match 3.6%; Score 19; DB 69; Length 287;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 ttctgtcaaaaaaa 526  
|||||  
Db 25 TTCTGTCAAAAAAAA 7

RESULT 13  
AA057492  
LOCUS AU057492 306 bp mRNA EST 29-APR-1999  
DEFINITION Oryza sativa mature leaf Nipponbare Oryza sativa cDNA  
clone S21487\_22, mRNA sequence.  
ACCESSION AU057492  
VERSION AU057492.1 GI:4716376  
KEYWORDS EST.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
AUTHORS euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
TITLE Poaceae; Oryza.  
JOURNAL 1 (bases 1 to 306)  
Yamamoto,K. and Sasaki,T.  
COMMENT Rice cDNA from mature leaf  
Unpublished (1999)  
On Jun 5, 1998 this sequence version replaced gi:3187742.  
Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai,Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: tsasaki@abr.affrc.go.jp  
PROJECT = RGP.

FEATURES  
source

Location/Qualifiers  
1..306  
/organism="Oryza sativa"  
/strain="Nipponbare"  
/db\_xref="taxon:4530"  
/clone\_lib="S21487\_22"  
/clone\_lib="Oryza sativa mature leaf Nipponbare"  
/tissue\_type="mature leaf"  
BASE COUNT 77 a 80 c 75 g 70 t 4 others  
ORIGIN

Query Match 3.6%; Score 19; DB 49; Length 306;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 ttctgtcaaaaaaa 526  
|||||  
Db 278 TTCTGTCAAAAAAAA 296

RESULT 14  
AA050528/c  
LOCUS AA050528 310 bp mRNA EST 09-SEP-1996  
DEFINITION mj19e04.r1 Soares mouse embryo NDME13.5 14.5 Mus musculus cDNA  
clone IMAGE:476574 5', mRNA sequence.  
ACCESSION AA050528  
VERSION AA050528.1 GI:1530394  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 310)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
On Oct 18, 1995 this sequence version replaced gi:1023851.  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:287318  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 279.

# FEATURES

source  
1. .310  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:476574"  
/clone\_lib="Soares mouse embryo NbME13.5 14.5"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"  
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCGAAGTGGCGCGCGGAAATTTTTTTTTTTTTTTTTT  
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2  
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne  
State Univ., from 2 ]; double-stranded cDNA was ligated to  
Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pT7T3 vector. Library went through one round of  
normalization, and was constructed by Bento Soares and  
M.Fatima Bonaldo."

BASE COUNT 88 a 69 c 68 g 85 t  
ORIGIN

Query Match 3.6%; Score 19; DB 27; Length 310;  
Best Local Similarity 100.0%; Pred. NO. 45;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 ttctctcaaaaaaaaaa 526  
|||||  
Db 24 TTCTGTCACAAAAA 6

## RESULT 15

C99144  
LOCUS C99144 317 bp mRNA EST 19-OCT-1998  
DEFINITION C99144 Rice panicle at ripening stage Oryza sativa cDNA clone  
E10258\_62, mRNA sequence.  
ACCESSION C99144  
VERSION C99144.1 GI:3761896  
KEYWORDS EST.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Oryza.  
1 (bases 1 to 317)  
Sasaki, T.  
Rice cDNA from panicle at ripening stage  
Unpublished (1996)  
JOURNAL On Jan 14, 1998 this sequence version replaced gi:1798792.  
COMMENT Contact: Takuji Sasaki

National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: tsasaki@abr.affrc.go.jp  
PROJECT "RGP"

## FEATURES

source  
1. .317  
Location/Qualifiers  
/organism="Oryza sativa"  
/strain="Nipponbare"  
/db\_xref="taxon:4530"  
/clone="E10258\_62"  
/clone\_lib="Rice panicle at ripening stage"  
/dev\_stage="ripening stage"  
/note="Organ: panicle; Rice cDNA from panicle at ripening  
stage"

BASE COUNT 87 a 61 c 88 g 76 t 5 others  
ORIGIN

Query Match 3.6%; Score 19; DB 43; Length 317;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 ttctgtcaaaaaaaaaa 526  
|||||  
Db 294 TTCTGTCACAAAAA 312

Search completed: May 1, 2000, 14:44:22  
Job time: 13982 sec



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:19:26 ; Search time 4425.31 Seconds  
(without alignments)  
-371.888 Million cell updates/sec

Title: US-09-215-435-78  
Perfect score: 542  
Sequence: 1 caccgacctgtggccatgat.....aaaaaaaaaaaaaaaaaaaaa 542

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0  
Searched: 821193 seqs, -1518192014 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pil.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vi.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vi.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_ba1.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*

- 44: gb\_htg6.\*
- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	30	5.5	171182	41	AC009266	AC009266 Homo sapi
c 2	29	5.4	93001	40	AC007286	AC007286 Homo sapi
c 3	29	5.4	194896	33	AC007457	AC007457 Mus muscu
c 4	29	5.4	217769	33	AC007438	AC007438 Mus muscu
c 5	28	5.2	181773	40	AC005881	AC005881 citb_79_e
c 6	28	5.2	235956	32	PFMALP2	AL035475 Plasmodiu
c 7	27	5.0	2467	4	AF005026	AF005026 Salvelinu
c 8	27	5.0	216693	43	AC013429	AC013429 Homo sapi
c 9	26	4.8	940	5	II5012	II5012 Sequence 1
c 10	26	4.8	940	5	I84551	I84551 Sequence 1
c 11	26	4.8	1192	10	HUMANTCD9	M38690 Human CD9 a
c 12	26	4.8	1881	12	AF099972	AF099972 Mus muscu
c 13	26	4.8	1984	8	AF015523	AF015523 Triticum
c 14	26	4.8	5545	7	VSEBPIGN	X95955 V.sativa mr
c 15	26	4.8	14005	35	AE001409	AE001409 Plasmodiu
c 16	26	4.8	69850	7	AB010700	AB010700 Arabidops
c 17	26	4.8	80250	44	AC015778	AC015778 Homo sapi
c 18	26	4.8	142519	11	HS620E11	AL031667 Human DNA
c 19	26	4.8	146705	40	AC007971	AC007971 Homo sapi
c 20	26	4.8	155187	32	HSJ540H1	AL121674 Homo sapi
c 21	26	4.8	256172	33	AC005139	AC005139 Plasmodiu
c 22	25	4.6	142	8	AF183873	AF183873 Abies fra
c 23	25	4.6	148	8	AF183872	AF183872 Abies fra
c 24	25	4.6	150	8	AF183871	AF183871 Abies bal
c 25	25	4.6	153	13	G19818	G19818 human STS A
c 26	25	4.6	405	34	TS0012669	AJ012669 Taenia so
c 27	25	4.6	437	10	HSNOBNDN3	U31338 Human nucle
c 28	25	4.6	488	35	AF010513	AF010513 Plasmodiu
c 29	25	4.6	518	10	HSM800832	AL110182 Homo sapi
c 30	25	4.6	533	8	AF061072	AF061072 Pneumocys
c 31	25	4.6	669	34	AB026847	AB026847 Ciona int
c 32	25	4.6	766	8	AF051369	AF051369 Oryza sat
c 33	25	4.6	786	7	OSRNACB5	X75670 O.sativa mr
c 34	25	4.6	889	12	HAMPLII	M27146 Hamster (M.
c 35	25	4.6	900	5	E03099	E03099 DNA encodin
c 36	25	4.6	900	5	E03166	E03166 Rat prolact
c 37	25	4.6	924	5	E03987	E03987 DNA encodin
c 38	25	4.6	945	7	S65090S2	S65092 lip2-lipase
c 39	25	4.6	959	10	S71381	S71381 prosome bet
c 40	25	4.6	1003	10	HSX15227	Y15227 Homo sapien
c 41	25	4.6	1020	10	HSU95301	U95301 Human calci
c 42	25	4.6	1053	4	AF102545	AF102545 Scaphiopu
c 43	25	4.6	1054	5	I08241	I08241 Sequence 1
c 44	25	4.6	1054	5	I08245	I08245 Sequence 6
c 45	25	4.6	1054	5	I08247	I08247 Sequence 9

ALIGNMENTS

RESULT 1  
AC009266/c  
LOCUS  
DEFINITION Homo sapiens chromosome 18 clone 481\_E\_14 map 18, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 6 unordered pieces.  
ACCESSION AC009266

11-AUG-1999

```

AC009266.1 GI:5729716
HTG: HTGS_PHASE1.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
1 (bases 1 to 171182)
TITLE Homo sapiens chromosome 18, clone 481_E_14
JOURNAL
REFERENCE
AUTHORS 2 (bases 1 to 171182)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Hearford,A., Horton,L., Howland,J.C., Jones,C., Kaul,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Nilloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (11-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2039: contig of 2039 bp in length
*
* 1
* 2040 4265: contig of 2226 bp in length
*
* 4266 14232: contig of 9967 bp in length
*
* 14233 27345: contig of 13113 bp in length
*
* 27346 131591: contig of 104246 bp in length
*
* 131592 171182: contig of 39591 bp in length.
FEATURES
source
Location/Qualifiers
1. .171182
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="481_E_14"
/clone_lib="RPCI-11 Human Male BAC"
BASE COUNT 55386 a 30342 c 29814 g 54471 t 1169 others
ORIGIN
Query Match 5.5% Score 30; DB 41; Length 171182;
Best Local Similarity 100.0%; Pred. No. 2.4e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 513 ttttcaatgaaaaaataaaaaaataaaaaa 542
Db 165190 TTTTCATGAAAAAATAAAAAAATAAAAAA 165161
RESULT 2

```

```

AC007286/c
LOCUS
DEFINITION Homo sapiens 12q15 BAC RPC111-298M11 (Roswell Park Cancer Institute
Human BAC Library) complete sequence.
ACCESSION AC007286
VERSION AC007286.18 GI:5819124
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 93001)
Muzny,D., Arenson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,
Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H.,
Gorrell,L.L., Hernandez,J., Issar,A., Jackson,L., Knetsch,S.,
Kondejowski,N., Lau,S., Leal,B., Lee,E., Lichtarge,O., Liu,W.,
Logan,O., Lu,J., Marodel,I., Martinez,C., Merscher,S., Miller,A.,
Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,
Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M.,
Vo,Q., Williamson,A., Worley,K.C., Xiang,A.M., Yang,R., Yu,W.,
Zhou,X., Kucherlapati,R., Nelson,D., Nickerson,E. and Gibbs,R.A.
Direct Submission
2 (bases 1 to 93001)
Worley,K.C.
Direct Submission
Submitted (11-APR-1999) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 93001)
Worley,K.C.
Direct Submission
Submitted (02-JUL-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 93001)
Worley,K.C.
Direct Submission
Submitted (02-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 93001)
Worley,K.C.
Direct Submission
Submitted (21-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 2, 1999 this sequence version replaced gi:5332370.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

```

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

#### ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2

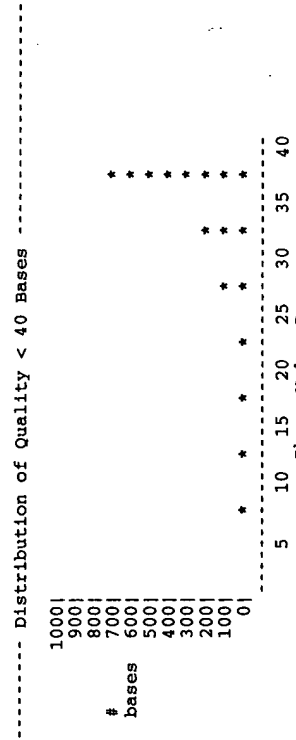
clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----  
Contig length: 93001  
Phrap values in estimate: 92957  
Average error rate (RCM-Phrap estimate): 7.45937e-05  
Fraction of Phrap values less than 40 : 0.0137268  
Number of consensus changing edits: 25  
Number of N's in consensus : 0

----- Consensus changing edits -----  
Position Original+Context Edited+Context  
9070 tgagttgatt(n)taagttcttgg tgagttgatt(c)taagttcttgg  
9084 gtctttgcta(n)tgtaacacgt gtctttgcta(t)tgtaacacgt  
18009 tctttctgg(n)ttttttttt tctttctgg(t)ttttttttt  
18010 ctctctgg(n)ttttttttt ctctctgg(t)ttttttttt  
18023 ttttttttt(n)ttttttttt ttttttttt(t)ttttttttt  
20975 ttttttttt(n)ttttttttt ttttttttt(t)ttttttttt  
34181 tgggaaggtta(n)agggagaggg tgggaaggtta(g)agggagaggg  
34224 ccacgcccgg(n)caatttgtta ccacgcccgg(c)caatttgtta  
34274 tccatgttgt(n)caagctggtc tccatgttgt(t)caagctggtc  
34289 cctcagcctc(n)caagttttg cctcagcctc(c)caagttttg  
34313 gttttggat(n)acagcgcatga gttttggat(c)acagcgcatga  
34315 actgcacctg(n)cnagtttgt actgcacctg(g)cnagtttgt  
34316 tgcaccctgn(n)agtttgttn tgcaccctgn(c)agtttgttn  
34316 gcaacctgn(n)agtttgttn gcaacctgn(c)agtttgttn  
34325 nagtttgt(n)naacatgga nagtttgt(t)naacatgga  
34326 nagtttgt(n)naacatgga nagtttgt(t)naacatgga  
34327 agtttgttn(n)aacatggaag agtttgttn(g)aacatggaag  
34342 tgggaagaca(n)cggtctggtc tgggaagaca(a)cggtctggtc  
34390 actttggag(n)cccgagaggg actttggag(g)cccgagaggg  
34460 agcccatct(n)taataaaaa agcccatct(c)taataaaaa  
35498 ggagctagca(n)aaagtatggt ggagctagca(g)aaagtatggt  
81693 aaaaatcagg(n)gtgtgttta aaaaatcagg(t)gtgtgttta  
81895 ggaagaaaag(n)accacacttg ggaagaaaag(a)accacacttg  
81979 atactaggtg(n)tgntattatg atactaggtg(t)tgntattatg  
81982 ctagggtgtg(n)tattatgaaa ctagggtgtg(t)tattatgaaa  
82322 tggtccttaa(n)aggggctggt tggtccttaa(g)aggggctggt



Version: 1.01 qfso.  
Location/Qualifiers

source  
1. .93001  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12q15"  
/clone="RPC111-298M11"  
1. .471  
/note="Overlaps bases 174297 to 174707 in AC007617."  
/function="Overlaps with adjacent clone AC007617."  
518. .631  
/rpt\_family="L2"  
1794. .1909  
/rpt\_family="L1M4"  
1998. .2102  
/rpt\_family="L1M4"  
complement(2103. .2403)  
/rpt\_family="AluY"  
2404. .2626  
/rpt\_family="L1M4"  
2729. .3168  
/rpt\_family="L1MEC"  
3169. .3460  
/rpt\_family="AluJb"  
3461. .3899  
/rpt\_family="L1MEC"  
complement(3900. .4072)  
/rpt\_family="FRAM"  
4073. .4268  
/rpt\_family="L1MEC"  
4312. .4529  
/rpt\_family="L1M4"  
complement(4530. .4896)  
/rpt\_family="MLT1A1"  
4897. .4952  
/rpt\_family="L1M4"  
complement(5463. .5759)  
/rpt\_family="AluSp"  
5972. .6090  
/rpt\_family="L2"  
6762. .6911  
/rpt\_family="L2"  
complement(6917. .7029)  
/rpt\_family="L2"  
7120. .7702  
/rpt\_family="MLT1F1"  
complement(7907. .8380)  
/rpt\_family="MER4B"  
complement(8425. .8627)  
/rpt\_family="MLT1A1"  
complement(8659. .15076)  
/rpt\_family="LIP3"  
complement(15089. .15353)  
/rpt\_family="MER4D"  
16022. .16310  
/rpt\_family="MLT1J"  
complement(16591. .17299)  
/rpt\_family="MLT1J"

Query Match 5.4%; Score 29; DB 40; Length 93001;  
Best Local Similarity 100.0%; Pred. No. 8.5e-05;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 514 ttctaatgaaaaaaaaaaaaaaaaaaaaa 542  
DB 55423 TTTCAATGAAAAAAAAAAAAAAAAAAAA 55395

RESULT 3  
AC007457 AC007457 194896 bp DNA HTG 03-JUN-1999  
LOCUS AC007457 Mus musculus, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
DEFINITION AC007457  
ACCESSION AC007457  
VERSION AC007457.10 GI:4966320  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE house mouse.

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS

Muzny, D., Arenson, A.D., Brundage, E., Carvelli, K., Chen, E., Di, W.,  
Ding, Y., Dugan, S., Durbin, J., Forcum, J., Ganesh, R., Garcia, C.,  
Goodman, M., Gorrell, J.H., Haywood, M., Jackson, L., Kampal, R.,  
Karpman, S., Leal, B., Li, Y., Liu, W., Logan, O., Lu, J., Ly, T.,  
Martinez, C., Oswal, G., Perez, L., Rashid, N.D., Rowland, K.,  
Savage, L., Scherer, S.S., Shen, H., Timms, K.M., Todd, J., Vo, Q.,  
Worley, K.C., Yu, W., Chinault, C., Nelson, D. and Gibbs, R.A.  
Direct Submission  
Unpublished  
2 (bases 1 to 194896)

TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Submitted (02-MAY-1999), Molecular and Human Genetics, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jun 2, 1999 this sequence version replaced gi:4955151.

## COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 50 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 43397: contig of 43397 bp in length  
\* 43398: gap of unknown length  
\* 43439: contig of 39854 bp in length  
\* 83293: gap of unknown length  
\* 83334: contig of 27534 bp in length  
\* 110868: gap of unknown length  
\* 110908: contig of 23050 bp in length  
\* 113059: gap of unknown length  
\* 133959: contig of 18344 bp in length  
\* 134000: gap of unknown length  
\* 152344: gap of unknown length  
\* 152385: contig of 1855 bp in length  
\* 154240: gap of unknown length  
\* 154281: gap of unknown length  
\* 155570: contig of 1289 bp in length  
\* 155611: gap of unknown length  
\* 156891: contig of 1281 bp in length  
\* 156892: gap of unknown length  
\* 156933: contig of 1248 bp in length  
\* 158181: gap of unknown length  
\* 158221: gap of unknown length  
\* 159457: contig of 1236 bp in length  
\* 159458: gap of unknown length  
\* 159499: contig of 1224 bp in length  
\* 160723: gap of unknown length  
\* 160763: contig of 1177 bp in length  
\* 161941: gap of unknown length  
\* 161981: contig of 1166 bp in length  
\* 163147: gap of unknown length  
\* 163148: gap of unknown length  
\* 163189: contig of 1125 bp in length  
\* 164314: gap of unknown length  
\* 164355: contig of 1106 bp in length  
\* 165461: gap of unknown length  
\* 165501: gap of unknown length  
\* 165502: contig of 1087 bp in length  
\* 166589: gap of unknown length  
\* 166630: contig of 1067 bp in length  
\* 167697: gap of unknown length  
\* 167737: contig of 1049 bp in length  
\* 168738: gap of unknown length  
\* 168827: contig of 1034 bp in length  
\* 168828: gap of unknown length  
\* 169802: contig of 1021 bp in length  
\* 169903: gap of unknown length  
\* 170924: contig of 1009 bp in length  
\* 170964: gap of unknown length  
\* 171973: contig of 983 bp in length  
\* 171974: gap of unknown length  
\* 172015: contig of 983 bp in length  
\* 172957: gap of unknown length  
\* 173038: contig of 767 bp in length  
\* 173059: gap of unknown length  
\* 173806: gap of unknown length

173847: contig of 924 bp in length  
\* 174771: gap of unknown length  
\* 174811: contig of 379 bp in length  
\* 175190: gap of unknown length  
\* 175191: gap of unknown length  
\* 175232: contig of 922 bp in length  
\* 176154: gap of unknown length  
\* 176195: contig of 896 bp in length  
\* 177091: gap of unknown length  
\* 177132: contig of 888 bp in length  
\* 178020: gap of unknown length  
\* 178061: contig of 866 bp in length  
\* 178927: gap of unknown length  
\* 178968: contig of 830 bp in length  
\* 179798: gap of unknown length  
\* 179838: contig of 818 bp in length  
\* 180657: gap of unknown length  
\* 180657: contig of 806 bp in length  
\* 181504: gap of unknown length  
\* 181544: contig of 622 bp in length  
\* 181545: gap of unknown length  
\* 182167: gap of unknown length  
\* 182207: contig of 781 bp in length  
\* 182208: gap of unknown length  
\* 183029: gap of unknown length  
\* 183030: contig of 731 bp in length  
\* 183761: gap of unknown length  
\* 183801: contig of 764 bp in length  
\* 183802: gap of unknown length  
\* 184566: contig of 763 bp in length  
\* 184607: gap of unknown length  
\* 185370: contig of 718 bp in length  
\* 185411: gap of unknown length  
\* 186129: contig of 716 bp in length  
\* 186170: gap of unknown length  
\* 186885: contig of 716 bp in length  
\* 186926: gap of unknown length  
\* 187641: contig of 716 bp in length  
\* 187642: gap of unknown length  
\* 187681: contig of 716 bp in length  
\* 188397: gap of unknown length  
\* 188437: contig of 714 bp in length  
\* 188438: gap of unknown length  
\* 189151: contig of 708 bp in length  
\* 189192: gap of unknown length  
\* 189900: contig of 705 bp in length  
\* 189939: gap of unknown length  
\* 190644: contig of 704 bp in length  
\* 190645: gap of unknown length  
\* 190685: contig of 700 bp in length  
\* 191389: gap of unknown length  
\* 191428: contig of 695 bp in length  
\* 192129: gap of unknown length  
\* 192169: contig of 691 bp in length  
\* 192864: gap of unknown length  
\* 192903: contig of 691 bp in length  
\* 193594: gap of unknown length  
\* 193595: contig of 690 bp in length  
\* 193635: gap of unknown length  
\* 194325: contig of 532 bp in length.  
\* 194365: Location/Qualifiers  
1.194896  
/organism="Mus musculus"  
/db\_xref="taxon:10090"

FEATURES  
source

BASE COUNT 50674 a 42623 c 43801 g 55745 t 2053 others  
ORIGIN

Query Match 5.4%; Score 29; DB 33; Length 194896;  
Best Local Similarity 100.0%; Pred. No. 7.9e-05;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 514 ttccatgaaaaaataaaaaa 542

Db 156121 TTTCATGAAAAA 156149

## RESULT 4

AC007438  
LOCUS AC007438 217769 bp DNA HTG 29-MAY-1999

```

DEFINITION Mus musculus, *** SEQUENCING IN PROGRESS ***, 59 unordered pieces.
ACCESSION AC007438
VERSION AC007438.6 GI:4926836
KEYWORDS HTG; HTGS_PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 217769)
Muzny,D., Aronson,A.D., Brundage,E., Carvelli,K., Chen,E., Di,W.,
Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R., Garcia,C.,
Goodman,M., Gorrell,J.H., Haywood,M., Jackson,L., Kampal,R.,
Karpathy,S., Leal,B., Li,Y., Liu,W., Logan,O., Lu,J., Ly,T.,
Martinez,C., Oswal,G., Perez,L., Rashid,N.D., Rowland,K.,
Savage,L., Scherer,S.S., Shen,H., Timms,K.M., Todd,J., Vo,Q.,
Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 217769)
Worley,K.C.
Direct Submission
Submitted (30-APR-1999) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 29, 1999 this sequence version replaced gi:4887716.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 59 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 33021: contig of 33021 bp in length
* 33022 33053: gap of unknown length
* 33054 62517: contig of 29464 bp in length
* 62518 62549: gap of unknown length
* 62550 85762: contig of 23213 bp in length
* 85763 85794: gap of unknown length
* 85795 105929: contig of 20135 bp in length
* 105930 105961: gap of unknown length
* 105962 121404: contig of 15443 bp in length
* 121405 121436: gap of unknown length
* 121437 132464: contig of 11028 bp in length
* 132465 132496: gap of unknown length
* 132497 140795: contig of 8299 bp in length
* 140796 140827: gap of unknown length
* 140828 147151: contig of 6324 bp in length
* 147152 147183: gap of unknown length
* 147184 151292: contig of 4109 bp in length
* 151293 151324: gap of unknown length
* 151325 155031: contig of 3707 bp in length
* 155032 155063: gap of unknown length
* 155064 158637: contig of 3574 bp in length
* 158638 158669: gap of unknown length
* 158670 161448: contig of 2779 bp in length
* 161449 161480: gap of unknown length
* 161481 163708: contig of 2228 bp in length
* 163709 163740: gap of unknown length
* 163741 165945: contig of 2205 bp in length
* 165946 165977: gap of unknown length
* 165978 167952: contig of 1975 bp in length
* 167953 167984: gap of unknown length
* 167985 169783: contig of 1799 bp in length
* 169784 169815: gap of unknown length
* 169816 171610: contig of 1795 bp in length
* 171611 171642: gap of unknown length
* 171643 173346: contig of 1704 bp in length
* 173347 173378: gap of unknown length
* 173379 174984: contig of 1606 bp in length
* 174985 175016: gap of unknown length
* 175017 176617: contig of 1601 bp in length
* 176618 176649: gap of unknown length
* 176650 178238: contig of 1589 bp in length
*
* 178239 178270: gap of unknown length
* 178271 179838: contig of 1568 bp in length
* 179839 179870: gap of unknown length
* 179871 181431: contig of 1561 bp in length
* 181432 181463: gap of unknown length
* 181464 182985: contig of 1522 bp in length
* 182986 183017: gap of unknown length
* 183018 184408: contig of 1391 bp in length
* 184409 184440: gap of unknown length
* 184441 185828: contig of 1388 bp in length
* 185829 185860: gap of unknown length
* 185861 187247: contig of 1355 bp in length
* 187248 187275: gap of unknown length
* 187276 188536: contig of 1289 bp in length
* 188537 188568: gap of unknown length
* 188569 189723: contig of 1155 bp in length
* 189724 189754: gap of unknown length
* 189755 190881: contig of 1127 bp in length
* 190882 190912: gap of unknown length
* 190913 192022: contig of 1110 bp in length
* 192023 192053: gap of unknown length
* 192054 193162: contig of 1109 bp in length
* 193163 193193: gap of unknown length
* 193194 194292: contig of 1099 bp in length
* 194293 194323: gap of unknown length
* 194324 195410: contig of 1087 bp in length
* 195411 195441: gap of unknown length
* 195442 196517: contig of 1076 bp in length
* 196518 196548: gap of unknown length
* 196549 197619: contig of 1071 bp in length
* 197620 197650: gap of unknown length
* 197651 198685: contig of 1035 bp in length
* 198686 198716: gap of unknown length
* 198717 199745: contig of 1029 bp in length
* 199746 199776: gap of unknown length
* 199777 200803: contig of 1027 bp in length
* 200804 200834: gap of unknown length
* 200835 203851: contig of 1017 bp in length
* 203852 203882: gap of unknown length
* 203883 203898: contig of 1016 bp in length
* 203899 203929: gap of unknown length
* 203930 203942: contig of 1013 bp in length
* 203943 203973: gap of unknown length
* 203974 204977: contig of 1004 bp in length
* 204978 205008: gap of unknown length
* 205009 206008: contig of 1000 bp in length
* 206009 206039: gap of unknown length
* 206040 206410: contig of 371 bp in length
* 206411 208441: gap of unknown length
* 208442 207412: contig of 971 bp in length
* 207413 207443: gap of unknown length
* 207444 208365: contig of 922 bp in length
* 208366 208396: gap of unknown length
* 208397 209287: contig of 891 bp in length
* 209288 209318: gap of unknown length
* 209319 210201: contig of 883 bp in length
* 210202 210232: gap of unknown length
* 210233 211115: contig of 883 bp in length
* 211116 211146: gap of unknown length
* 211147 212027: contig of 881 bp in length
* 212028 212058: gap of unknown length
* 212059 212937: contig of 879 bp in length
* 212938 212968: gap of unknown length
* 212969 213769: contig of 801 bp in length
* 213770 213800: gap of unknown length
* 213801 214586: contig of 786 bp in length
* 214587 214617: gap of unknown length
* 214618 215398: contig of 781 bp in length
* 215399 215429: gap of unknown length
* 215430 216173: contig of 744 bp in length
* 216174 216204: gap of unknown length
* 216205 218864: contig of 680 bp in length
* 218865 218895: gap of unknown length

```

\* 216896 217521: contig of 626 bp in length  
 \* 217522 217552: gap of unknown length  
 \* 217553 217769: contig of 217 bp in length.

## FEATURES

source  
 1. 217769  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"

BASE COUNT 56451 a 51343 c 52045 g 55795 t 2135 others  
 ORIGIN

Query Match 5.4%; Score 29; DB 33; Length 217769;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-05;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 514 ttcaatgaaaaa542  
 Db 172563 TTTCAATGAAAAA542

## RESULT 5

LOCUS AC005881 181773 bp DNA PRI 12-NOV-1999  
 DEFINITION citb-79\_e16, complete sequence.

ACCESSION AC005881  
 VERSION AC005881.3 GI:6382477  
 KEYWORDS HTG.

## SOURCE

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS Smith, D.R.  
 TITLE Sequencing of Human Chromosome 10  
 JOURNAL Unpublished

## REFERENCE

AUTHORS Smith, D.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-OCT-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA

## REFERENCE

AUTHORS Smith, D.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA

## REFERENCE

AUTHORS Smith, D.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-MAR-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA

## REFERENCE

AUTHORS Smith, D.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA

## REMARK

COMMENT On Nov 12, 1999 this sequence version replaced gi:4314322.

## FEATURES

source  
 1. 181773  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /map="10q25"  
 /clone\_lib="CIT987SK-1079E16"  
 2 others

## BASE COUNT

ORIGIN 53635 a 35777 c 37389 g 54970 t  
 Query Match 5.2%; Score 28; DB 40; Length 181773;  
 Best Local Similarity 100.0%; Pred. No. 0.00026;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 515 ttcaatgaaaaa542

Db 107328 TTCAATGAAAAA107301

## RESULT 6

PFMAL4P2/c

LOCUS

DEFINITION

PROGRESS \*\*\*

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Aug 12, 1999 this sequence version replaced gi:5531347.

For more information about this sequence or the Malasia Project.

see http://www.sanger.ac.uk/projects/P-falciparum. IMPORTANT: This

sequence is unfinished and does not necessarily represent the

correct sequence. Work on the sequence is in progress and the

release of this data is based on the understanding that the

sequence may change as work continues. The sequence may be

contaminated with foreign sequence from E.coli, yeast, vector,

phage etc.

Order of segments is not known; 800 n's separate segments.

\* NOTE: This is a 'working draft' sequence.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

FEATURES

Location/Qualifiers

1. 235956

/organism="Plasmodium falciparum"

/strain="3D7"

/db\_xref="taxon:5833"

/chromosome="4"

BASE COUNT 90265 a 25047 c 22408 g 97435 t 801 others

ORIGIN

Query Match 5.2%; Score 28; DB 32; Length 235956;

Best Local Similarity 100.0%; Pred. No. 0.00026;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 515 ttcaatgaaaaa542

Db 129848 TTCAATGAAAAA129821

## RESULT 7

AF005026

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Query Match

Best Local Similarity

Matches

QY 515

Db 129848

TTCAATGAAAAA129821

AF005026

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Query Match

Best Local Similarity

Matches

QY 515

Db 129848

TTCAATGAAAAA129821

AF005026

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Query Match

Best Local Similarity

Matches

QY 515

Db 129848

TTCAATGAAAAA129821

AF005026

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Query Match

Best Local Similarity

Matches

QY 515

Db 129848

TTCAATGAAAAA129821

AF005026

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Query Match

Best Local Similarity

Matches

QY 515

Db 129848

TTCAATGAAAAA129821

AF005026

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Query Match

Best Local Similarity

Matches

QY 515

Db 129848

TTCAATGAAAAA129821

AF005026

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Query Match

Best Local Similarity

Matches

QY 515

Db 129848

TTCAATGAAAAA129821

AF005026

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Query Match

Best Local Similarity

Matches

QY 515

Db 129848

TTCAATGAAAAA129821

AF005026

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Query Match

Best Local Similarity

Matches

QY 515

Db 129848

TTCAATGAAAAA129821

AF005026

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Query Match

Best Local Similarity

Matches

QY 515

Db 129848

TTCAATGAAAAA129821

AF005026

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Query Match

Best Local Similarity

Matches

QY 515

Db 129848

TTCAATGAAAAA129821

AF005026

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM



```
the effects of kallikrein-like serine proteases on follicle
contraction
Biol. Reprod. 58 (4), 887-897 (1998)
98206794
REFERENCE
2 (bases 1 to 2467)
Hajnik,C.A., Goetz,F.W. and Hsu,S.-Y.
Direct Submission
Submitted (21-MAY-1997) Biological Sciences, University of Notre
Dame, Notre Dame, IN 46556, USA
FEATURES
Location/Qualifiers
1..2467
/organism="Salvelinus fontinalis"
/db_xref="taxon:8038"
/tissue_type="preovulatory ovarian tissue"
/clone="KT14"
712..1479
CDS
/codon_start=1
/product="serine protease-like protein precursor"
/protein_id="AAC17927.1"
/db_xref="GI:2286143"
/translation="MGERRFLTAALIALGFLCSDAITGGREAAHSPYPYMASLQVAD
GDRKHGGGLVADQWVMSAAHCFLSGEGKVKVLGAHSLSEPEDSKOTFDIVQVTS
HPDFSINVDIALIKLDRPIMASDAVKSILQSDGADPATQEVNTAGWGLNLI
GSRDQLQELVIDVMNRVRCGRSDYGGKFTNNMLCAASRORDICDGDGGPLLYKGV
AVGITSNGKKGCKGSKKPLGYTISHYSQWIDKTMQ"
712..777
sig_peptide
mat_peptide
817..1476
/product="serine protease-like protein"
polyA_site 2442
BASE COUNT 601 c 527 g 572 t
ORIGIN

Query Match 5.0%; Score 27; DB 4; Length 2467;
Best Local Similarity 100.08; Pred. No. 0.0013;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 tcaatgaaaaaataaaaaaataaaaaa 542
|||||
Db 2436 TCAATGAAAAAATAAAAAAATAAAAAA 2462

RESULT 8
AC013429/c
LOCUS
DEFINITION Homo sapiens chromosome 7, *** SEQUENCING IN PROGRESS ***, 90
unordered pieces.
ACCESSION AC013429
VERSION AC013429.5 GI:6453918
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 216693)
Olson,M.V.
Direct Submission
TITLE
Unpublished
REFERENCE
2 (bases 1 to 216693)
Bubb,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M.
Direct Submission
Submitted (10-NOV-1999) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
On Nov 19, 1999 this sequence version replaced gi:6449515.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
```

```
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 90 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1038: contig of 1038 bp in length
* gap of unknown length
* 1039 2245: contig of 1207 bp in length
* gap of unknown length
* 2246 3404: contig of 1159 bp in length
* gap of unknown length
* 3405 4432: contig of 1028 bp in length
* gap of unknown length
* 4433 5534: contig of 1102 bp in length
* gap of unknown length
* 5535 5943: contig of 409 bp in length
* gap of unknown length
* 5944 6924: contig of 981 bp in length
* gap of unknown length
* 6925 7981: contig of 1057 bp in length
* gap of unknown length
* 7982 8964: contig of 983 bp in length
* gap of unknown length
* 8965 9194: contig of 230 bp in length
* gap of unknown length
* 9195 10187: contig of 993 bp in length
* gap of unknown length
* 10188 10689: contig of 502 bp in length
* gap of unknown length
* 10690 11322: contig of 633 bp in length
* gap of unknown length
* 11323 12710: contig of 1388 bp in length
* gap of unknown length
* 12711 13157: contig of 447 bp in length
* gap of unknown length
* 13158 14440: contig of 1283 bp in length
* gap of unknown length
* 14441 15698: contig of 1258 bp in length
* gap of unknown length
* 15699 16715: contig of 1017 bp in length
* gap of unknown length
* 16716 16814: contig of 99 bp in length
* gap of unknown length
* 16815 17804: contig of 990 bp in length
* gap of unknown length
* 17805 19512: contig of 1708 bp in length
* gap of unknown length
* 19513 19934: contig of 422 bp in length
* gap of unknown length
* 19935 20397: contig of 463 bp in length
* gap of unknown length
* 20398 21539: contig of 1142 bp in length
* gap of unknown length
* 21540 22709: contig of 1170 bp in length
* gap of unknown length
* 22710 23848: contig of 1139 bp in length
* gap of unknown length
* 23849 24058: contig of 210 bp in length
* gap of unknown length
* 24059 25768: contig of 1710 bp in length
* gap of unknown length
* 25769 26476: contig of 708 bp in length
* gap of unknown length
* 26477 27382: contig of 906 bp in length
* gap of unknown length
* 27383 29078: contig of 1696 bp in length
* gap of unknown length
* 29079 30000: contig of 922 bp in length
* gap of unknown length
```

```

* 30001 30851: contig of 851 bp in length
*      gap of unknown length
* 30852 31018: contig of 167 bp in length
*      gap of unknown length
* 31019 32296: contig of 1278 bp in length
*      gap of unknown length
* 32297 33283: contig of 987 bp in length
*      gap of unknown length
* 33284 33529: contig of 246 bp in length
*      gap of unknown length
* 33530 34041: contig of 512 bp in length
*      gap of unknown length
* 34042 34917: contig of 876 bp in length
*      gap of unknown length
* 34918 36145: contig of 1228 bp in length
*      gap of unknown length
* 36146 37782: contig of 1637 bp in length
*      gap of unknown length
* 37783 39451: contig of 1669 bp in length
*      gap of unknown length
* 39452 41109: contig of 1638 bp in length
*      gap of unknown length
* 41110 42143: contig of 1034 bp in length
*      gap of unknown length
* 42144 43080: contig of 937 bp in length
*      gap of unknown length
* 43081 45006: contig of 1926 bp in length
*      gap of unknown length
* 45007 45472: contig of 466 bp in length
*      gap of unknown length
* 45473 46997: contig of 1525 bp in length
*      gap of unknown length
* 46998 47564: contig of 567 bp in length
*      gap of unknown length
* 47565 48568: contig of 1004 bp in length
*      gap of unknown length
* 48569 50128: contig of 1560 bp in length
*      gap of unknown length
* 50129 50561: contig of 433 bp in length
*      gap of unknown length
* 50562 51525: contig of 964 bp in length
*      gap of unknown length
* 51526 53243: contig of 1718 bp in length
*      gap of unknown length
* 53244 54826: contig of 1583 bp in length
*      gap of unknown length
* 54827 56175: contig of 1349 bp in length
*      gap of unknown length
* 56176 56344: contig of 169 bp in length
*      gap of unknown length
* 56345 57820: contig of 1476 bp in length
*      gap of unknown length
* 57821 58776: contig of 956 bp in length
*      gap of unknown length
* 58777 59857: contig of 1081 bp in length
*      gap of unknown length
* 59858 62122: contig of 2265 bp in length
*      gap of unknown length
* 62123 63568: contig of 1446 bp in length
*      gap of unknown length
* 63569 65693: contig of 2125 bp in length
*      gap of unknown length
* 65694 66788: contig of 1095 bp in length
*      gap of unknown length
* 66789 68875: contig of 2087 bp in length
*      gap of unknown length
* 68876 70344: contig of 1469 bp in length
*      gap of unknown length
* 70345 73019: contig of 2675 bp in length
*      gap of unknown length
* 73020 76543: contig of 3524 bp in length
*      gap of unknown length
* 76544 79986: contig of 3443 bp in length

```

```

* 79987 80928: contig of 942 bp in length
*      gap of unknown length
* 80929 85088: contig of 4160 bp in length
*      gap of unknown length
* 85089 88323: contig of 3235 bp in length
*      gap of unknown length
* 88324 91351: contig of 3028 bp in length
*      gap of unknown length
* 91352 95305: contig of 3954 bp in length
*      gap of unknown length
* 95306 99154: contig of 3849 bp in length
*      gap of unknown length
* 99155 103098: contig of 3944 bp in length
*      gap of unknown length
* 103099 108849: contig of 5751 bp in length
*      gap of unknown length
* 108850 112623: contig of 3774 bp in length
*      gap of unknown length
* 112624 116499: contig of 3876 bp in length
*      gap of unknown length
* 116500 120998: contig of 4499 bp in length
*      gap of unknown length
* 120999 126802: contig of 5804 bp in length
*      gap of unknown length

```

```

Query Match      5.0%; Score 27; DB 43; Length 216693;
Best Local Similarity 100.0%; Pred. No. 0.00085;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 516 tcaatgagaaaaa542
Db 71334 TCAATGAGAAAAA71308

```

```

RESULT 9
LOCUS I15012 940 bp DNA PAT 02-APR-1996
DEFINITION Sequence 1 from patent US 5457178.
ACCESSION I15012
VERSION I15012.1 GI:1249920
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 940)
AUTHORS Jackson,J.R.H., Krapcho,K.J., Johnson,J.H. and Kral,R.M. Jr.
TITLE Insecticidally effective spider toxin
JOURNAL Patent: US 5457178-A 1 10-OCT-1995;
FEATURES
source
BASE COUNT 322 a 152 c 193 g 259 t 14 others
ORIGIN

```

```

Query Match      4.8%; Score 26; DB 5; Length 940;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 517 caatgagaaaaa542
Db 907 CAATGAGAAAAA932

```

```

RESULT 10
LOCUS I84551 940 bp DNA PAT 04-APR-1998
DEFINITION Sequence 1 from patent US 5695959.
ACCESSION I84551
VERSION I84551.1 GI:3022071
KEYWORDS
SOURCE Unknown.

```

```

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 940)
AUTHORS Jackson,J.RandolphHunter, Krapcho,K.Joanne, Johnson,J.Helen and Kral,R.Marden Jr.
TITLE Recombinant expression of insecticidally effective spider toxin
JOURNAL Patent: US 5695959-A 1 09-DEC-1997;
FEATURES
    Location/Qualifiers
        source
            1..940
                /organism="unknown"
BASE COUNT 322 a 152 c 193 g 259 t 14 others
ORIGIN

Query Match 4.8%; Score 26; DB 5; Length 940;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 517 caatgaaaaa542
Db 907 CAATGAAAAA932

RESULT 11
LOCUS HUMANTCD9 1192 bp mRNA PRI 01-NOV-1995
DEFINITION Human CD9 antigen mRNA, complete cds.
ACCESSION M38690
VERSION M38690.1 GI:1048988
KEYWORDS CD9 antigen.
SOURCE Human cDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1192)
AUTHORS Boucheix,C., Benoit,P., Frachet,P., Billard,M., Worthington,R.E., Gagnon,J., and Uzan,G.
TITLE Molecular cloning of the CD9 antigen. A new family of cell surface proteins
JOURNAL J. Biol. Chem. 266 (1), 117-122 (1991)
MEDLINE 91093112
REFERENCE 2 (bases 1 to 1192)
AUTHORS Boucheix,C.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1990) Claude Boucheix, Unite 268, INSERM, Av. Paul Vaillant-Couturier, Villejuif, 94800, France
COMMENT On Nov 1, 1995 this sequence version replaced gi:178674.
FEATURES
    Location/Qualifiers
        1..1192
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_type="megacaryocyte"
            /map="12p13"
            /chromosome="12"
            /tissue_type="hemopoietic system"
            52..738
                /gene="CD9"
            52..738
                /gene="CD9"
            /codon_start=1
            /db_xref="GDB:G00-120-582"
            /product="CD9 antigen"
            /protein_id="AAA80320.1"
            /db_xref="GI:1048989"
            /translation="MPVKGKTKYLLFGNFIFWLAGIYAVLGLWLRDPSOTKSI
            FEQTNNNNSFYGVYILIGAGLMLVGLGCGGAVQESQCMGLGFEGELLVIFAI
            EIRAAWGYSHKDEIVKEVQFYKDTNKLKATKDEPREILKAHYALNCCGLAGGVE
            QFISDPCPDLVETFTVKSCPDAIKEVFDNKFHILIGAVGIGIAVWIFGMIFMILC
            CAIRNRKEMV"
BASE COUNT 310 a 243 c 273 g 366 t
ORIGIN

```

```

Query Match 4.8%; Score 26; DB 10; Length 1192;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 517 caatgaaaaa542
Db 1163 CAATGAAAAA1188

RESULT 12
LOCUS AF099972 1881 bp mRNA ROD 09-DEC-1998
DEFINITION Mus musculus schlafeni (Sifnl) mRNA, complete cds.
ACCESSION AF099972
VERSION AF099972.1 GI:3983147
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1881)
AUTHORS Schwarz,D.A., Katayama,C.D. and Hedrick,S.M.
TITLE Schlafen, a new family of growth regulatory genes that affect thymocyte development
JOURNAL Immunity 9 (5), 657-668 (1998)
MEDLINE 99060921
REFERENCE 2 (bases 1 to 1881)
AUTHORS Schwarz,D.A., Katayama,C.D. and Hedrick,S.M.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-1998) Cancer Center and Biology Department, UC San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0687
FEATURES
    Location/Qualifiers
        1..1881
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="11"
            /map="49 cM"
            1..1881
                /gene="slfn1"
                258..1271
                    /gene="slfn1"
                    /note="mslfn1"
                    /codon_start=1
                    /product="schlafen1"
                    /protein_id="AAC83825.1"
                    /db_xref="GI:3983148"
                    /translation="MNITDEGTVLILNAGGITIGTSRKTMENHVRNRTTKAL
                    CALINGSGEKVKAHIKNPDYILSKHGIGEDLETSFKNILSRPLDFKQYOSYFICVE
                    KSQSPDVSVGKPAITIANLNMKGASVVENLDAQKFLDNIKVAGRSFARSDDRP
                    GDDTQEEGHIOELAAFFKQSKLTKEFLFSKSNVEYKFSFKLLQVRKELLPT
                    VSAFANTDGGYLFGLDEKKEIVGFEAKNCQPKLESEIKCIQQLPVTFFCEERK
                    IKYKCFIEVHDSGVCKYVCALRVFCCAVFAAPESWHMKDGGVKRFTIEWIKL
                    LMS"
BASE COUNT 554 a 423 c 427 g 477 t
ORIGIN

Query Match 4.8%; Score 26; DB 12; Length 1881;
Best Local Similarity 100.0%; Pred. No. 0.0043;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 517 caatgaaaaa542
Db 1853 CAATGAAAAA1878

RESULT 13
LOCUS AF015523 1984 bp mRNA PLN 29-SEP-1999
DEFINITION Triticum aestivum low-affinity cation transporter (LCT1) mRNA, complete cds.
ACCESSION AF015523
VERSION AF015523.1 GI:2460043

```





us-09-215-435-78.olig.rge

Fri May 12 12:28:02 2000

-----

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 18:51:03 ; Search time 250.34 Seconds  
(without alignments)  
541.680 Million cell updates/sec

Title: US-09-215-435-78  
Perfect score: 542  
Sequence: 1 cagcagctgtggccatgat.....aaaaaaaaaaaaaaaaaaaaa 542

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 311585 seqs, 125096042 residues

Word size : 0

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	228	42.1	421	1 X51465	Human secreted pro
2	28	5.2	1905	1 T33552	Glutathione synth
3	26	4.8	226	1 V60636	5'RACE product fro
4	26	4.8	940	1 Q8268	Spider venom FIL-7
5	26	4.8	1982	1 V10366	Wheat low affinity
6	26	4.8	3056	1 X26149	Human pan-s/tk-1B
7	26	4.8	4467	1 X05820	Rat pheromone rece
8	25	4.6	43	1 V48090	Oligonucleotide 43
9	25	4.6	545	1 X20418	Human secreted pro
10	25	4.6	637	1 X22221	Human secreted pro
11	25	4.6	882	1 T72173	Alzheimer's diseas
12	25	4.6	899	1 Q14451	Rat prolactin gene
13	25	4.6	924	1 Q31989	Guinea pig MBP-2 g
14	25	4.6	1057	1 Q04285	CDNA clone of sequ
15	25	4.6	1147	1 X00658	Human secreted pro
16	25	4.6	1290	1 Q15023	Rat prolactin codi
17	25	4.6	1508	1 T37312	Aromatic acyl tran
18	25	4.6	1537	1 T72662	Human smooth muscl
19	25	4.6	1872	1 Q90623	Murine osteogenic
20	25	4.6	1873	1 T33442	Mouse osteogenic p
21	25	4.6	1873	1 Q24517	Murine osteogenic
22	25	4.6	1873	1 Q28736	Murine osteogenic
23	25	4.6	1873	1 Q38945	Mouse osteogenic p
24	25	4.6	1873	1 Q38734	Murine pro-OP-1. M
25	25	4.6	1873	1 Q38858	Morphogen mOP1 cod
26	25	4.6	1873	1 Q53153	Sequence encoding
27	25	4.6	1873	1 Q58199	mOP1 cDNA. Morphog
28	25	4.6	1873	1 Q58232	mOP1 cDNA. Use mor
29	25	4.6	1873	1 Q58051	Mouse osteogenic p
30	25	4.6	1873	1 Q67312	Murine OP-1. Morph
31	25	4.6	1873	1 Q45117	Murine OP-1. Maint
32	25	4.6	1873	1 Q65392	Osteogenic protein
33	25	4.6	1873	1 Q45163	Murine OP-1. Use o

34	25	4.6	1873	1 Q72704	mOP1-PP prepro for
35	25	4.6	1873	1 T02598	mOP-1 cDNA. Antibo
36	25	4.6	1873	1 T97879	cDNA encoding mous
37	25	4.6	1873	1 V10346	Mouse Op-1 cDNA. T
38	25	4.6	1873	1 V15216	Mouse osteogenic p
39	25	4.6	1873	1 V19534	Mouse osteogenic p
40	25	4.6	1873	1 V32584	Mouse osteogenic p
41	25	4.6	1873	1 X00241	Murine osteogenic
42	25	4.6	1969	1 V34250	Human secreted pro
43	25	4.6	2042	1 X00621	Human secreted pro
44	25	4.6	2309	1 Q73786	Partial PCR fragme
45	25	4.6	2524	1 T18673	Human trophinin CD

## ALIGNMENTS

RESULT 1	
X51465	
ID X51465;	
AC X51465;	
DE 21-JUN-1999 (first entry)	
DT Human secreted protein 5', EST SEQ ID NO:44.	
DE Human; secreted protein; EST; expressed sequence tag; diagnosis;	
KW forensic; gene therapy; chromosome mapping; signal peptide;	
KW upstream regulatory sequence; cytokine activity; cell proliferation;	
KW differentiation; haematopoiesis regulation; tissue growth regulation;	
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;	
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.	
OS Homo sapiens.	
PN W0906349-A2.	
PD 11-FEB-1999.	
PF 31-JUL-1998; IB1231.	
PR 01-AUG-1997; US-905279.	
PA (GEST ) GENSET.	
PI Duclert A. Dumas Milne Edwards J, Lacroix B;	
DR WPI: 99-153779/13.	
DR P-PSDB: Y12687.	
PT New nucleic acids encoding human secreted proteins - obtained from	
PT CDNA libraries derived from testis, ovary, uterus and spleen tissue	
PS Claim 1: Page 177; 522pp; English.	
CC X51459 to X51691 represent 5' expressed sequence tags (ESTs) for human	
CC secreted proteins, and encode the proteins given in Y12681 to Y12913,	
CC respectively. The proteins given represent the signal peptide and an	
CC N-terminal fragment of a secreted protein. The nucleic acid sequences	
CC can be used for producing secreted human gene products. They can also	
CC be used to develop products for diagnosis and therapy. The proteins	
CC obtained may have cytokine activity, cell proliferation/differentiation	
CC activity, haematopoiesis regulating activity, tissue growth regulating	
CC activity, reproductive hormone regulating activity, chemotactic/	
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/	
CC ligand activity, anti-inflammatory activity, tumour inhibition activity	
CC or other activities. The products can be used in forensic, gene therapy	
CC and chromosome mapping procedures. The sequences can also be used for	
CC obtaining corresponding promoter sequences. The nucleic acids encoding	
CC the signal peptide can be used for directing extracellular secretion of	
CC a polypeptide or the insertion of a polypeptide into a membrane, or	
CC importing a polypeptide into a cell.	
SQ Sequence 421 BP; 96 A; 106 C; 110 G; 102 T;	

Query Match	42.1%	Score 228;	DB 1;	Length 421;
Best Local Similarity	100.0%;	Pred. No. 1.8e-79;		
Matches 228;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 137	ggaaccagagactgcgagactgctgctgccaagctgtctccagacaatttcgagtcacat	196
DB 183	GGAAACAGAGACTGCGAGACTGGCTGCTGCCAACGTCTCCAGACAATTTCGAGTCGCACT	242
QY 197	gcgcggagagaagggtccgagggcagctctgtctcaacgcaggtgtcttcttggccaata	256
DB 243	GGCGGAGAGAGGGGTCGCCAGGCGAGTCTGTGTCAACAGCAGGTGTCTTTGGCCATA	302

V60636:  
23-NOV-1998 (first entry)  
5'RACE product from paralysis tick HT-I cDNA.  
Neurotoxin; HT-I; Australian paralysis tick; Ixodes holocyclus; PCR;  
primer; RACE; rapid amplification of cDNA ends; tryptic digest;  
immune response; cow; horse; goat; cat; dog; ss.  
Synthetic.  
Ixodes holocyclus.  
OS Key Location/Qualifiers  
FH Key 1..60  
FT CDS /\*tag= a  
/\*note= "C-terminus of HT-I protein"  
WO9747649-A1.  
18-DEC-1997.  
PD 11-JUN-1997; AU0366.  
PF 11-JUN-1996; AU-000395.  
PR (INSE-) INSEARCH LTD.  
PA Broady KW, Masina S, Thurn MJ;  
PI WPI: 98-052243/05.  
DR P-PSDB; W68452.  
PT Isolated nucleic acid encoding paralysis tick neurotoxin - used to  
immunise animals against the neurotoxin, specifically that of Ixodes  
holocyclus  
PT Examples; Fig 11; 45pp; English.  
PS This sequence represents the product generated by a 5'-RACE reaction  
using primer V60634, which based on the sequence of PCR fragment (V60633)  
of the neurotoxin HT-I gene from the Australian paralysis tick - Ixodes  
holocyclus. When sequenced, the product revealed some differences  
between the amino acid sequence encoded by the cDNA and the sequence  
derived by amino acid sequencing the tryptic digest fragments. The new  
sequence was used to design new primers for obtaining the complete HT-I  
gene. The polypeptide or nucleic acid encoding it can be used to raise  
a protective immune response against paralytic tick neurotoxin,  
especially in cows, horses, goats, cats and dogs.  
SQ Sequence 226 BP; 105 A; 32 C; 48 G; 41 T;

Query Match 4.8%; Score 26; DB 1; Length 226;  
Best Local Similarity 100.0%; Pred. No. 0.1; Mismatches 0; Indels 0; Gaps 0;  
Matches 26; Conservative 0;

QY 517 caatgaataaaataaaataaaataaa 542  
|||||  
DB 195 CATGAATAAAATAAAATAAAATAAA 220  
|||||

RESULT 4  
Q85268 ID Q85268 standard; DNA; 940 BP.  
AC Q85268;  
DT 18-AUG-1995 (first entry)  
DE Spider venom FIL-705.  
KW Spider venom; FIL-376; FIL-705; insecticide; pesticide;  
KW Biological control agent; baculovirus; Lepidoptera; ds.  
OS Filistrata hibernalis.  
FH Key Location/Qualifiers  
FT misc\_difference 1..14  
FT /\*tag= a  
/\*note= "bases at positions 1-14 are not given in  
the specification"  
FT 1..750  
FT /\*tag= b  
WO9501996-A.  
PD 19-JAN-1995.  
PF 07-JUL-1994; U07595.  
PR 07-JUL-1993; US-089998.  
PA (FMCC) FMC CORP.  
PA (NPSP-) NPS PHARM INC.  
PI Jackson JRH, Kral RM, Krapcho KJ;  
DR WPI: 95-066867/09.  
DR P-PSDB; R70761.  
PT New insecticidal proteins from spider venom - and related  
expression vectors and transformed hosts, partic. for control of

QY 257 gacgtgtccctcgcgcgaactgactgtatattcaagaatgagaatgctta 316  
|||||  
DB 303 GAGCGTCCCTCCCTCGCGAAGCTGACTGTATATTCAAGAATGAGAATGCTTA 362  
|||||

QY 317 gcatcgctatgcccgtgtgcagaaattggaagcgagagtggtgcta 364  
|||||  
DB 363 GCATCGCCTATGCGCGTGTGCAGAAATTCGAGGCGAGAGTTGGCTA 410  
|||||

RESULT 2  
T33552 ID T33552 standard; DNA; 1905 BP.  
AC T33552;  
DT 18-APR-1997 (first entry)  
DE Glutathione synthetase coding sequence.  
KW Glutathione synthetase; GSH; rat; kidney; glutathione; metabolism; RFLP;  
KW cellular protection; transhydrogenation; coenzyme A; human; antibody;  
KW 5-oxoprolinuria; metabolic acidosis; gene therapy; probe; PCR detection;  
KW restriction fragment length polymorphism analysis; therapy; ss.  
OS Rattus rattus.  
FH Key Location/Qualifiers  
FT CDS 45..1469  
FT /\*tag= a  
/\*product= glutathione synthetase  
WO9625172-A1.  
22-AUG-1996.  
PD 08-FEB-1996; U01660.  
PF 13-FEB-1995; US-387117.  
PR (CORR) CORNELL RES FOUND INC.  
PA Anderson ME, Huang C, Meister A;  
PI WPI: 96-393135/39.  
DR P-PSDB; W00669.  
PT New recombinant rat kidney glutathione synthase and related DNA -  
useful in gene therapy of glutathione deficiency and for synthesis  
of glutathione analogues  
PS Claim 2; Page 35-36; 51pp; English.  
CC This sequence represents the coding sequence for the glutathione  
synthetase (GSH) isolated from rat kidney. Glutathione functions in  
metabolism, transport, and cellular protection. Specifically,  
glutathione participates in transhydrogenation reactions that are  
involved in the formation and maintenance of the sulphhydryl group of  
other molecules, e.g. coenzyme A. GSH deficiency in humans, has two  
general types, both of which lead to serious health complications. In  
one type, an unstable form of GSH is expressed, leading to an apparently  
selective deficiency of GSH in erythrocytes. In the other type, such as  
5-oxoprolinuria, potentially fatal metabolic consequences occur as a  
result of over-production of 5-oxoprolinone, leading to severe metabolic  
acidosis. Introduction of this sequence into a cell will increase its  
glutathione content, e.g. for treating mammals suffering from glutathione  
deficiency (e.g. 5-oxoprolinuria), and in patients undergoing radiation  
or chemo-therapy, and to protect against toxic side effects of drugs.  
The encoded protein can be used to prepare radiolabelled glutathione  
derivatives, and to raise antibodies for use in immunoassays. Fragments  
of this sequence can be used as probes to detect GSH in body fluids, in  
PCR detection processes, or in restriction fragment length polymorphism  
analysis. Antisense nucleic acid molecules targeting this sequence can  
be used to reduce glutathione production, which may be useful when  
administering chemotherapeutic agents that are detoxified by glutathione.  
SQ Sequence 1905 BP; 474 A; 499 C; 530 G; 402 T;

Query Match 5.2%; Score 28; DB 1; Length 1905;  
Best Local Similarity 100.0%; Pred. No. 0.014; Mismatches 0; Indels 0; Gaps 0;  
Matches 28; Conservative 0;

QY 515 ttcaatgaataaaataaaataaaataaa 542  
|||||  
DB 1876 TTCATGAATAAAATAAAATAAAATAAA 1903  
|||||

RESULT 3  
V60636 ID V60636 standard; cDNA; 226 BP.



```

PT Lepidoptera larvae.
PS Disclosure; Page 42-43; 64pp; English.
CC Oligonucleotides corresp. to the partial sequence of insecticidal
CC protein FIL-376 were used as primers for PCR amplification of
CC Filistrata hibernalis cephalothorax cDNA. Products were
CC inserted into pKS and used to transform E. coli. Subclone
CC FIL-705 encoded a 249-amino acid protein (given in R70761) that
CC may be a precursor of FIL-376. Expression of such proteins e.g.
CC from baculovirus vectors is used for biological control of
CC Lepidopteran pests.
SQ Sequence 940 BP; 322 A; 152 C; 193 G; 259 T;

Query Match 4.8%; Score 26; DB 1; Length 940;
Best Local Similarity 100.0%; Pred. No. 0.087;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 517 caatgaaaaaataaaaaa 542
Db 907 CAATGAAAAAATAAAAAA 932

RESULT 5
ID V10366 standard; cDNA; 1982 BP.
AC V10366;
DT 01-JUL-1998 (first entry)
DE Wheat low affinity cation transporter LCT1 gene.
KW Wheat; low affinity cation transporter; LCT1; heavy metal; alkali;
OS Triticum sp.
FH Key Location/Qualifiers
FT CDS 137..1474
FT /*tag= a
FT /product= "LCT1"
PN WO9804700-A1.
PD 05-FEB-1998.
PF 28-JUL-1997; U13250.
PR 29-JUL-1996; US-022722.
PA (REGC ) UNIV CALIFORNIA.
PI Antosiewicz DM, Clemens S, Schachtman DP, Schroeder JJ;
DR WPI; 98-130689/12.
DR P-PSDB; W40204.
PT Plant low affinity cation transporter - used for modulating heavy or
PT alkali metal uptake
PS Claim 1; Page 24-25; 31pp; English.
CC The present sequence encodes wheat low affinity cation transporter
CC (LCT1). The present invention also describes a transgenic plant,
CC preferably a member of the genus Brassica, comprising an expression
CC cassette containing a plant promoter operably linked to a heterologous
CC LCT1 polynucleotide. The LCT1 polynucleotide can be used in an
CC expression cassette for modulating heavy metal or alkali metal uptake
CC in a plant. The transgenic plants are useful for removing heavy metals
CC from contaminated soils.
SQ Sequence 1982 BP; 354 A; 611 C; 569 G; 448 T;

Query Match 4.8%; Score 26; DB 1; Length 1982;
Best Local Similarity 100.0%; Pred. No. 0.08;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 517 caatgaaaaaataaaaaa 542
Db 1955 CAATGAAAAAATAAAAAA 1980

RESULT 6
ID X26149 standard; cDNA; 3056 BP.
AC X26149;
DT 21-MAY-1999 (first entry)
DE Human pan-s/tk-1B receptor polypeptide encoding cDNA.
KW Pancreatic serine/threonine kinase receptor; pan-s/tk; cell growth;

```

```

KW differentiation; gene therapy; ligand; screening; cell proliferation;
KW regulator; pancreatic; ds.
PN OS Homo sapiens.
PN WO9907854-A2.
PD 18-FEB-1999.
PF 11-AUG-1998; U16640.
PR 11-AUG-1997; US-909354.
PA (ONTO-) ONTOGENY INC.
PI Barker DD, Miao N, Pang K;
DR WPI; 99-167430/14.
DR P-PSDB; Y01497.
PT New pancreatic serine/threonine kinase (pan-s/tk) receptor
PT polypeptides - useful for modulating cell proliferation,
PT differentiation and survival in animals
PS Claim 28; Page 68-69; 77pp; English.
CC The invention relates to isolated and/or recombinant pancreatic serine/
CC theonine kinase receptor (pan-s/tk) polypeptides. Sequences of rat and
CC human pan-s/tk polypeptides and nucleic acids encoding them are provided.
CC Host cells containing vectors comprising the pan-s/tk genes are
CC used for the recombinant expression of the proteins. The pan-s/tk
CC polypeptides are useful for modulating cell growth, differentiation or
CC survival in an animal cell. The polynucleotides are also useful for
CC detecting pan-s/tk ligand on cells in samples and for gene therapy.
CC Ligands identified, together with soluble polypeptides, are useful for
CC screening test compounds that modulate the bioactivity of a pan-s/tk
CC receptor. The discovery of this new class of pan-s/tk polypeptides
CC provides an alternative method of mediating cell proliferation,
CC differentiation and survival in animals. The polypeptides also represent
CC the first identified specific regulators of pancreatic development. The
CC present sequence represents a cDNA encoding a human pan-s/tk-1B receptor
CC polypeptide.
SQ Sequence 3056 BP; 953 A; 591 C; 645 G; 867 T;

Query Match 4.8%; Score 26; DB 1; Length 3056;
Best Local Similarity 100.0%; Pred. No. 0.077;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 517 caatgaaaaaataaaaaa 542
Db 3031 CAATGAAAAAATAAAAAA 3056

RESULT 7
ID X05820 standard; cDNA; 4467 BP.
AC X05820;
DT 04-MAY-1999 (first entry)
DE Rat pheromone receptor Go-VN4 encoding cDNA.
KW Pheromone receptor; signal transduction; fertility; behaviour;
KW reproduction; rodent; insect; rat; ss.
OS Rattus sp.
FH Key Location/Qualifiers
FT CDS 126..2726
FT /*tag= a
PN WO9900422-A1.
PD 07-JAN-1999.
PF 30-JUN-1998; U13680.
PR 30-JUN-1997; US-051284.
PA (HARD ) HARVARD COLLEGE.
PI Buck L, Dulac C, Herrada G, Matsunami H;
DR WPI; 99-095684/08.
DR P-PSDB; W94920.
PT New isolated pheromone receptor polypeptides - used to develop
PT products for controlling fertility and behaviour in vertebrates and
PT invertebrates
PS Claim 18; Page 136-140; 308pp; English.
CC The invention relates to polynucleotide sequences encoding mammalian
CC pheromone receptor polypeptides. The polypeptides are expressed in
CC murine and rat vomeronasal organ. The products can be used for modifying
CC pheromone activity, e.g. for decreasing pheromone receptor mediated
CC signal transduction. They can be used for controlling fertility and
CC behaviour in vertebrates and invertebrates. Compositions comprising the

```

CC polypeptides are particularly useful in e.g. controlling fertility in  
 CC livestock and controlling reproduction in rodents or insects by  
 CC interrupting the normal behaviours of rodents or insects that result in  
 CC reproduction. The present sequence represents a cDNA encoding a rat  
 CC pheromone receptor Go-VN4. The cDNA is deposited under the Genbank  
 CC accession number AF016181.  
 SQ Sequence 4467 BP; 1414 A; 918 C; 805 G; 1330 T;

Query Match 4.8%; Score 26; DB 1; Length 4467;  
 Best Local Similarity 100.0%; Pred. No. 0.074; 0; Indels 0; Gaps 0;  
 Matches 26; Conservative 0; Mismatches 0;

QY 517 caatgaaataaaataaaataaaataaa 542  
 |||:|||||  
 Db 4440 CAAAGAAAAAATAAATAAATAAATAA 4465

RESULT 8  
 V48090 ID V48090 standard; DNA; 43 BP.  
 AC V48090;  
 DT 27-OCT-1998 (first entry)  
 DE Oligonucleotide 43-P.  
 KW In situ translation; RNA-protein fusion; binding reagent; antibody;  
 KW Industrial catalyst; ss; rRNA/DNA hybrid.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT modified\_base 43  
 FT /\*tag= a  
 FT /\*note= "Puromycin"  
 FT  
 PN W09831700-A1.  
 PD 23-JUL-1998.  
 PE 14-JAN-1998; U00807.  
 PR 06-NOV-1997; US-064491.  
 PR 21-JAN-1997; US-035963.  
 PA (GHO) GEN HOSPITAL CORP.  
 PI Liu R, Roberts RW, Szostak JW;  
 DR WPI; 98-414032/35.  
 PT Selection of specific protein by screening protein-RNA fusions  
 PT generated in vitro or in situ - useful for, e.g. identifying enzymes  
 PT and antibodies with altered properties, potentially useful as  
 PT catalysts or for therapy or diagnosis  
 PS Disclosure; Page 39; 94pp; English.  
 CC The Oligonucleotides V48087, V48089-V48091 and V48096-V48098 and  
 CC variations were used to generate RNA-protein fusions. These were used in  
 CC for identification of a specific protein or RNA, by in vitro or in situ  
 CC translation of candidate RNA molecules to produce RNA-protein fusions,  
 CC then selecting specific RNA protein fusions. The method is used to  
 CC select proteins (or DNA encoding them) having altered properties, e.g.  
 CC for identification of new binding reagents, to identify improved human  
 CC antibodies or new enzymes. These proteins are potentially useful in  
 CC diagnosis and therapy, or as industrial catalysts. The methods allow  
 CC many rounds of selection and amplification to be performed, resulting in  
 CC enrichment of even very rare molecules and allowing isolation of proteins  
 CC that bind specifically to almost any compound or catalyse almost any  
 CC reaction.  
 SQ Sequence 43 BP; 32 A; 3 C; 6 G; 1 U;

Query Match 4.6%; Score 25; DB 1; Length 43;  
 Best Local Similarity 96.0%; Pred. No. 0.29; 0; Indels 0; Gaps 0;  
 Matches 24; Conservative 1; Mismatches 0;

QY 518 aatgaaataaaataaaataaaataaa 542  
 |||:|||||  
 Db 10 AAGAAAAAATAAATAAATAAATAA 34

RESULT 9  
 X20418 ID X20418 standard; DNA; 545 BP.  
 AC X20418;

DT 04-MAY-1999 (first entry)  
 DE Human secreted protein gene 7.  
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; lymphoma;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 OS Homo sapiens.  
 PN W09906423-A1.  
 PD 11-FEB-1999.  
 PE 29-JUL-1998; U15949.  
 PR 19-AUG-1997; US-056730.  
 PR 30-JUL-1997; US-054209.  
 PR 30-JUL-1997; US-054211.  
 PR 30-JUL-1997; US-054212.  
 PR 30-JUL-1997; US-054213.  
 PR 30-JUL-1997; US-054214.  
 PR 30-JUL-1997; US-054215.  
 PR 30-JUL-1997; US-054217.  
 PR 30-JUL-1997; US-054218.  
 PR 30-JUL-1997; US-054234.  
 PR 30-JUL-1997; US-054236.  
 PR 18-AUG-1997; US-053968.  
 PR 18-AUG-1997; US-053969.  
 PR 18-AUG-1997; US-053972.  
 PR 19-AUG-1997; US-056534.  
 PR 19-AUG-1997; US-056543.  
 PR 19-AUG-1997; US-056554.  
 PR 19-AUG-1997; US-056561.  
 PR 19-AUG-1997; US-056727.  
 PR 19-AUG-1997; US-056729.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Carter KC, Endress GA, Fan P, Feng P, Kiyaw H, Lafleur DW,  
 PI Li Y, Moore PA, Rosen CA, Ruben SM, Shi Y, Wei Y,  
 PI Zeng Z;  
 DR WPI; 99-153691/13.  
 DR P-PSDB; Y00264.  
 PT New isolated human genes and the secreted polypeptides they encode -  
 PT useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 PS Claim 1; Page 214; 312pp; English.  
 CC This sequence represents a nucleic acid molecule which encodes a secreted  
 CC human protein. The gene number is given in the descriptor line. The gene  
 CC can be used to generate fusion proteins by linking to the gene to a human  
 CC immunoglobulin Fc portion (e.g. X20403) for increasing the stability of  
 CC the fused protein as compared to the human protein only.  
 CC The invention relates to 83 novel genes and their fragments (nucleic acid  
 CC sequences: X20412-X20499; amino acid sequences Y00258-Y00377) which  
 CC are useful for preventing, treating or ameliorating medical conditions  
 CC e.g. by protein or gene therapy. Also, pathological conditions can be  
 CC diagnosed by determining the amount of the new polypeptides in a sample  
 CC or by determining the presence of mutations in the new polynucleotides.  
 CC Specific uses are described for each of the 86 polynucleotides, based on  
 CC which tissues they are most highly expressed in (see X20412 for described  
 CC uses).  
 SQ Sequence 545 BP; 137 A; 143 C; 164 G; 97 T;

Query Match 4.6%; Score 25; DB 1; Length 545;  
 Best Local Similarity 100.0%; Pred. No. 0.22; 0; Indels 0; Gaps 0;  
 Matches 25; Conservative 0; Mismatches 0;

QY 518 aatgaaataaaataaaataaaataaa 542  
 |||:|||||  
 Db 503 AATGAAAAAATAAATAAATAAATAA 527

RESULT 10  
 X22221 ID X22221 standard; DNA; 637 BP.

X22221; (first entry)  
AC 18-MAY-1999 Human secreted protein gene 11 clone HWT8K81.  
DE Human; secreted protein; gene therapy; protein therapy; cancer; weight;  
KW tumour; chromosome mapping; forensic; haematological disease; allergy;  
KW inflammation; cell proliferation; viral infection; wound healing;  
KW modulation; appetite; behaviour; food additive; preservative; ss.  
OS Homo sapiens.  
PN WO903990-A1.  
PD 28-JAN-1999.  
PF 15-JUL-1998; U14613.  
PR 18-AUG-1997; US-056361.  
PR 16-JUL-1997; US-052661.  
PR 16-JUL-1997; US-052870.  
PR 16-JUL-1997; US-052871.  
PR 16-JUL-1997; US-052872.  
PR 16-JUL-1997; US-052873.  
PR 16-JUL-1997; US-052874.  
PR 16-JUL-1997; US-052875.  
PR 22-JUL-1997; US-053440.  
PR 22-JUL-1997; US-053441.  
PR 22-JUL-1997; US-053442.  
PR 18-AUG-1997; US-055683.  
PR 18-AUG-1997; US-055724.  
PR 18-AUG-1997; US-055725.  
PR 18-AUG-1997; US-055726.  
PR 18-AUG-1997; US-055946.  
PR 18-AUG-1997; US-055952.  
PR 18-AUG-1997; US-055985.  
PR 18-AUG-1997; US-055989.  
PR 18-AUG-1997; US-056359.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Duan R, Feng P, Ferrie AM, Florence KA, Fouad J, Yu G, Greene JM, Hu J, Ni J, Rosen CA, Ruben SM, Young PE, Yu G.  
PI WPI: 99-132234/11.  
DR P-PSDB; Y01393.  
PT New nucleic acids encoding secreted human proteins - potentially useful for treating and diagnosing diseases and identifying specific binding agents  
PT Claim 4; Page 175; 251pp; English.  
PS The invention relates to nucleic acid sequences (X22211 to X22282) encoding human secreted proteins (Y01383 to Y01454). The secreted protein gene sequences are deposited with the ATCC under deposit number ATCC 209138, 209139 or 209141. Host cells containing vectors comprising the nucleic acid sequences are used for the recombinant expression of the secreted proteins. The polynucleotide and amino acid sequences are useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. pathological conditions can be also diagnosed by determining the amount of the new polypeptides in a sample or by the presence of mutations in the new polynucleotides. The nucleic acid sequences, or its fragments, are useful for chromosome identification and mapping; as antisense and triplex-forming therapeutics; in gene therapy; for (forensic) identification of individuals; as molecular weight markers; to identify related sequences or specific mRNA; in preparation of oligonucleotides and to raise anti-DNA antibodies. Antibodies are useful as immunoassay reagents (including for in vivo imaging) and therapeutically to inhibit or activate particular polypeptides. A very wide range of disorders may be treated with the polynucleotide and polypeptide sequences, e.g. autoimmune or haematological diseases, allergy, inflammation, cancer or other forms of cell proliferation, viral or other infections. The sequences may also be useful in wound healing, to modulate differentiation of embryonic stem cells, to modulate weight, appetite, behaviour etc. and as food additive or preservative. The present sequence represents a gene encoding a human secreted protein (see descriptor line for gene number and clone identification).  
CC Sequence 537 BP; 164 A; 164 C; 159 G; 149 T;  
SQ

Query Match 4.6%; Score 25; DB 1; Length 637;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 aatgaaaaaataaaaaa 542  
DB 609 AATGAAAAAATAAAAAA 633  
RESULT 11  
T72173  
ID T72173 standard; cDNA to mRNA; 882 BP.  
AC T72173;  
DT 25-FEB-1998 (first entry)  
DE Alzheimer's disease DNA sequence from plasmid pGCS1180.  
KW Human; brain; Alzheimer's disease; diagnosis; antibody; expression; ds.  
OS Homo sapiens.  
FH Key  
FT Location/Qualifiers  
FT 97..198  
FT /\*tag= a  
FT /\*note= "Protein encoded by DNA from the human brain which has different expression from the normal protein when in an Alzheimer's disease patient"  
PN WO9721807-A1.  
PD 19-JUN-1997.  
PF 12-DEC-1996; J03630.  
PR 12-DEC-1995; JP-322745.  
PA (KYOW) KYOWA HAKKO KOGYO KK.  
PI Hashida H, Kuga T, Nakagawa S, Sakaki Y, Zhao N;  
PI WPI: 97-332779/30.  
DR P-PSDB; W19950.  
PT DNA sequences with varied expression in normal and Alzheimer patients - used for diagnosis and treatment of Alzheimer's disease  
PT Claim 3; Page 48-49; 90pp; Japanese.  
PS The present sequence represents a DNA fragment which is complementary to mRNA sequences whose level of expression in the brains of Alzheimer's disease patients varies from normal individuals. The present sequence represents a specifically claimed DNA fragment in which the expression is augmented to at least twice that in normal cases. The sequence, and antibodies to the proteins encoded by the sequence, may be used in the diagnosis, treatment and investigation of Alzheimer's disease.  
CC Sequence 882 BP; 306 A; 134 C; 300 T;  
SQ

Query Match 4.6%; Score 25; DB 1; Length 882;  
Best Local Similarity 100.0%; Pred. No. 0.21;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 aatgaaaaaataaaaaa 542  
DB 826 AATGAAAAAATAAAAAA 850

RESULT 12  
Q14451  
ID Q14451 standard; DNA; 899 BP.  
AC Q14451;  
DT 21-JAN-1992 (first entry)  
DE Rat prolactin gene.  
KW Recombinant DNA; ss.  
OS Rattus rattus.  
FH Key  
FT Location/Qualifiers  
FT 52..732  
FT /\*tag= a  
PN J03219876-A.  
PD 27-SEP-1991.  
PF 24-JAN-1990; 014511.  
PR 24-JAN-1990; JP-014511.  
PA (SHIK-) SHIKISHIMA BOSEKI K.  
DR WPI: 91-329111/45.  
DR P-PSDB; R14599.

PT Mass-prodn. of rat prolactin-producing recombinant DNA - is by integration of promoter, Shine-Dalgarno sequence and translation initiation codon upstream of table 1 coding gene.  
PT Claim 1; Table 1; 12pp; Japanese.  
CC The gene can be ligated into an expression plasmid with a promoter, SD sequence and initiation codon for the prodn. of recombinant rat

CC prolactin. The protein is useful for the study of the physiological  
 CC activity of rat and human prolactin.  
 CC See also Q14452.  
 SQ Sequence 899 BP; 277 A; 206 C; 175 G; 241 T;

Query Match 4.6%; Score 25; DB 1; Length 899;  
 Best Local Similarity 100.0%; Pred. No. 0.21;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 518 aatgaataaaataaaataaaataaa 542  
 |||||  
 Db 867 AATGAATAAAATAAAATAAAATAAA 891

## RESULT 13

Q31989  
 ID Q31989 standard; cDNA; 924 BP.  
 AC Q31989;  
 DT 30-APR-1993 (first entry)  
 DE Guinea pig MBP-2 gene.  
 KW Major basic protein; anti-asthma; ss.  
 OS Caria porcellus.

Location/Qualifiers  
 FH Key 89..790  
 FT cds  
 FT signal\_peptide 89..134  
 FT /\*tag= a  
 FT signal\_peptide 135..433  
 FT /\*tag= b  
 FT signal\_peptide 135..433  
 FT /\*tag= c  
 FT /\*note= "precursor peptide"  
 FT mat\_peptide 434..790  
 FT /\*tag= d

PN J04320686-A.  
 PD 11-NOV-1992.  
 PF 22-APR-1991; 090704.  
 PR 22-APR-1991; JP-090704.  
 PA (SAKA ) ORSUKA PHARM CO LTD.  
 DR WPI; 92-426680/52.  
 DR P-PSDB; R29717.  
 PT A guinea pig MBP gene - contains DNA base sequence coding MBP  
 PT gene having 1 of 2 specified aminoacid sequences  
 PS Claim 1; Page 16; 20pp; Japanese  
 CC The sequence shows the cDNA sequence of the guinea pig major basic  
 CC protein (MBP)-2 gene. The sequence was obtd. by first isolating MBP-1  
 CC and MBP-2 from the abdominal transudate of male guinea pigs treated  
 CC with polymyxin and ascaris extract. MBP-1 and MBP-2 were used to  
 CC design probes for screening of a guinea pig eosinophil cDNA library.  
 CC The MBPs may be prep'd. in large amounts by this method and may be  
 CC used for the confirmation of the effect of an anti-asthma drug.  
 CC See also Q31988-93.  
 SQ Sequence 924 BP; 216 A; 236 C; 287 G; 185 T;

Query Match 4.6%; Score 25; DB 1; Length 924;  
 Best Local Similarity 100.0%; Pred. No. 0.21;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 518 aatgaataaaataaaataaaataaa 542  
 |||||  
 Db 896 AATGAATAAAATAAAATAAAATAAA 920

## RESULT 14

Q04285  
 ID Q04285 standard; cDNA; 1057 BP.  
 AC Q04285;  
 DT 17-SEP-1990 (first entry)

DE cDNA clone of sequence encoding bovine trophoblast protein-1 (bTP-1).  
 KW Bovine trophoblast protein-1; bTP-1; fertility; ds.  
 OS Sus scrofa.  
 FH Key Location/Qualifiers  
 FT cds 84..667

FT EP-367063-A. /\*tag= a

PN 9-MAY-1990.  
 PD 23-OCT-1989; 119642.  
 PR 26-OCT-1988; US-262870.  
 PA (UMOR) Univ of Missouri.  
 PI Roberts MR, Imakawa K;  
 DR WPI; 90-141062/19.  
 DR P-PSDB; R04539.

PT Recombinant bovine trophoblast protein-1 -  
 PT used for enhancing fertility or treating viral diseases in  
 PT mammals, esp. cattle.  
 PS Claim 3; Fig 1; 27pp; English.  
 CC The bTP-1 produced from the gene may be used to promote fertility  
 CC or treat viral disease in cattle. The gene may also be used to  
 CC provide transgenic animals with enhanced fertility, or in  
 CC prophylactic and therapeutic treatment of other mammals.  
 SQ Sequence 1057 BP; 294 A; 260 C; 229 G; 274 T;

Query Match 4.6%; Score 25; DB 1; Length 1057;  
 Best Local Similarity 100.0%; Pred. No. 0.21;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 518 aatgaataaaataaaataaaataaa 542  
 |||||  
 Db 1033 AATGAATAAAATAAAATAAAATAAA 1057

## RESULT 15

X00658  
 ID X00658 standard; DNA; 1147 BP.

AC X00658;  
 DT 25-MAR-1999 (first entry)  
 DE Human secreted protein gene 48 clone HTEJN13.  
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
 KW immune system; asthma; lymphocytic disease; brain; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 OS Homo sapiens.  
 PN W09842738-A1.

PD 01-OCT-1998.  
 PR 19-MAR-1998; U05311.  
 PR 30-MAY-1997; US-050937.  
 PR 21-MAR-1997; US-041276.  
 PR 21-MAR-1997; US-041277.  
 PR 21-MAR-1997; US-041281.  
 PR 21-MAR-1997; US-042344.  
 PR 30-MAY-1997; US-048069.  
 PR 30-MAY-1997; US-048094.  
 PR 30-MAY-1997; US-048095.  
 PR 30-MAY-1997; US-048096.  
 PR 30-MAY-1997; US-048099.  
 PR 30-MAY-1997; US-048131.  
 PR 30-MAY-1997; US-048135.  
 PR 30-MAY-1997; US-048154.  
 PR 30-MAY-1997; US-048160.  
 PR 30-MAY-1997; US-048186.  
 PR 30-MAY-1997; US-048187.  
 PR 30-MAY-1997; US-048188.  
 PR 30-MAY-1997; US-048350.  
 PR 30-MAY-1997; US-048351.  
 PR 30-MAY-1997; US-048352.  
 PR 30-MAY-1997; US-048355.  
 PR 05-AUG-1997; US-054804.

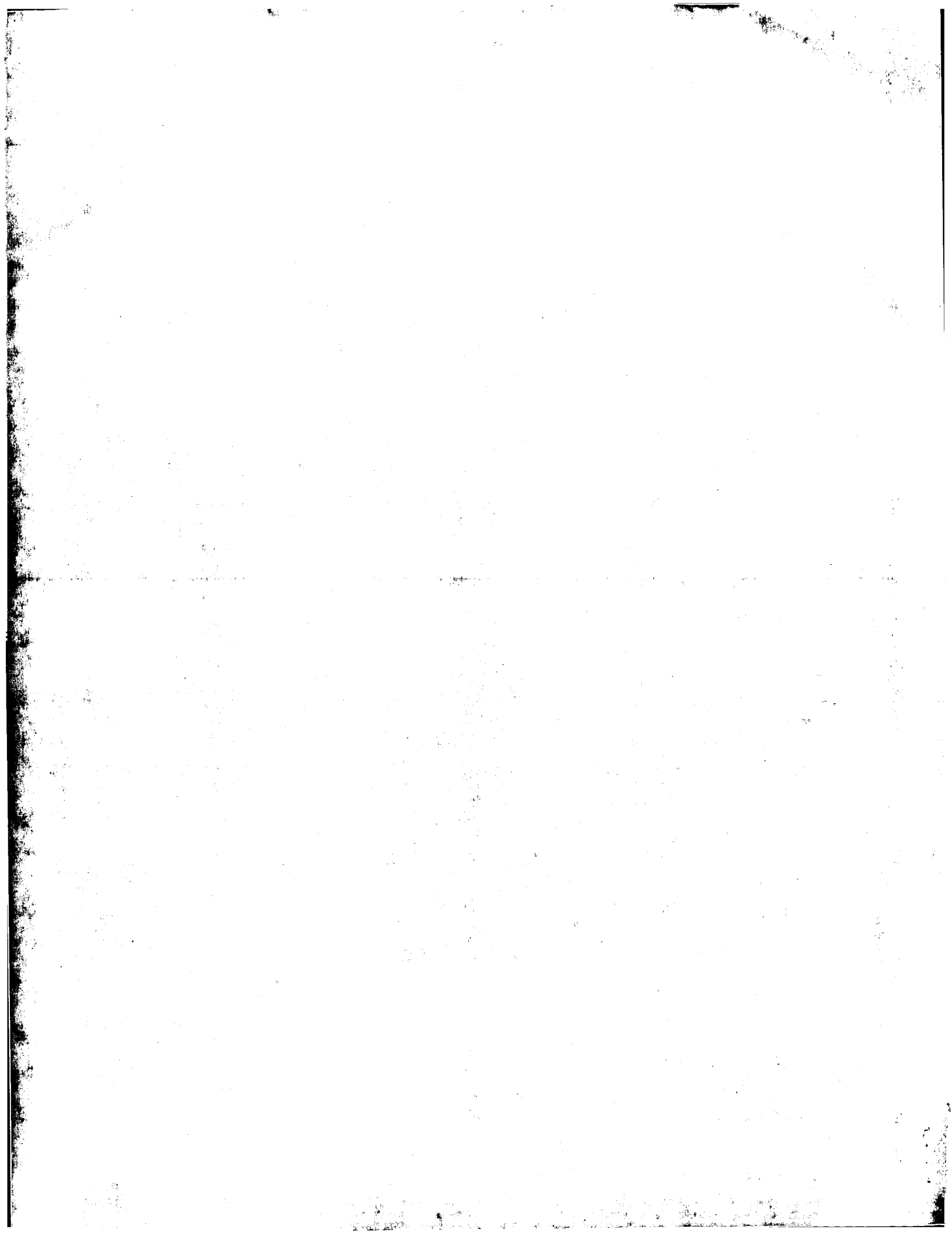
(HUMA-) HUMAN GENOME SCI INC.  
 PA Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA.  
 PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS,  
 PI Rosen CA, Ruben SM, Shi Y, Young P;

DR WPI: 99-070066/06.  
 DR P-PSDB; W67854.  
 PT New isolated human genes and the secreted polypeptides they encode -  
 PT useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 PS Claim 1: Page 210-211: 385pp; English.  
 CC This sequence represents a nucleic acid molecule which encodes a secreted  
 CC human protein. The gene number, and the clone it is derived from, are  
 CC detailed in the descriptor line. The gene can be used to generate fusion  
 CC proteins by linking to the gene to a human immunoglobulin Fc portion  
 CC (e.g. X00602) for increasing the stability of the fused protein as  
 CC compared to the human protein only.  
 CC The invention relates to 87 novel genes and their fragments (nucleic acid  
 CC sequences: X00611-X00724; amino acid sequences W67807-W68004) which  
 CC are useful for preventing, treating or ameliorating medical conditions  
 CC e.g. by protein or gene therapy. Also, pathological conditions can be  
 CC diagnosed by determining the amount of the new polypeptides in a sample  
 CC or by determining the presence of mutations in the new polynucleotides.  
 CC Specific uses are described for each of the 87 polynucleotides, based on  
 CC which tissues they are most highly expressed in (see X00611 for described  
 CC uses).  
 SQ Sequence 1147 BP; 338 A; 275 C; 264 G; 270 T;

Query Match 4.6%; Score 25; DB 1; Length 1147;  
 Best Local Similarity 100.0%; Pred. No. 0.21;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 aatgataaaaaaa  
 DB 1120 ATGTA

Search completed: May 1, 2000, 18:51:09  
 Job time: 18753 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:04:38 ; Search time 181.17 Seconds  
(without alignments)  
358.172 Million cell updates/sec

Title: US-09-215-435-78  
Perfect score: 542  
Sequence: 1 cagacactgtggccatgat.....aaaaaaaaaaaaaaaaaaaaa 542

Scoring table:  
Gapop 60.0 , Gapext 60.0

Searched: 214294 seqs, 59861574 residues

Word size : 0

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PCUS9\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	4.8	940	1	US-08-089-998B-1
2	26	4.8	940	2	US-08-457-272-1
3	26	4.8	940	6	PCT-US94-07595-1
4	26	4.6	1982	4	US-08-900-148-1
5	25	4.6	861	4	US-08-924-759-15
6	25	4.6	882	4	US-08-909-965C-9
7	25	4.6	1537	4	US-08-839-008-1
8	25	4.6	1873	1	US-07-841-646-24
9	25	4.6	1873	1	US-07-901-703-8
10	25	4.6	1873	1	US-08-147-023-24
11	25	4.6	1873	1	US-08-278-729A-18
12	25	4.6	1873	1	US-08-480-528A-5
13	25	4.6	1873	1	US-08-479-666-5
14	25	4.6	1873	1	US-08-155-343A-18
15	25	4.6	1873	1	US-08-408-672-18
16	25	4.6	1873	2	US-08-643-563A-18
17	25	4.6	1873	2	US-08-447-570-24
18	25	4.6	1873	2	US-08-643-763A-18
19	25	4.6	1873	2	US-08-462-623-18
20	25	4.6	1873	2	US-08-451-953A-18
21	25	4.6	1873	3	US-08-445-346-3
22	25	4.6	1873	3	US-08-445-468A-18
23	25	4.6	1873	3	US-08-901-200A-5
24	25	4.6	1873	3	US-08-449-700-24
25	25	4.6	1873	4	US-08-449-699A-24
26	25	4.6	1873	4	US-08-461-397A-18

27	25	4.6	1873	4	US-08-912-088-18	Sequence 18, Appl
28	25	4.6	1873	6	PCT-US90-07654-1	Sequence 1, Appl
29	25	4.6	1873	6	PCT-US92-01968-18	Sequence 18, Appl
30	25	4.6	1873	6	PCT-US93-05446-8	Sequence 8, Appl
31	25	4.6	1873	6	PCT-US93-07189-3	Sequence 3, Appl
32	25	4.6	1873	6	PCT-US93-07190-18	Sequence 18, Appl
33	25	4.6	1873	6	PCT-US93-07231-18	Sequence 18, Appl
34	25	4.6	1873	6	PCT-US93-08742-18	Sequence 18, Appl
35	25	4.6	1873	6	PCT-US93-08808-18	Sequence 18, Appl
36	25	4.6	1873	6	PCT-US93-08885-18	Sequence 18, Appl
37	25	4.6	1873	6	PCT-US93-10520-5	Sequence 5, Appl
38	25	4.6	2173	1	US-08-036-210-6	Sequence 6, Appl
39	25	4.6	2173	4	US-08-449-609-6	Sequence 6, Appl
40	25	4.6	2309	1	US-08-036-210-10	Sequence 10, Appl
41	25	4.6	2309	4	US-08-449-609-10	Sequence 10, Appl
42	25	4.6	2524	1	US-08-317-522A-1	Sequence 1, Appl
43	25	4.6	2524	1	US-08-439-818A-1	Sequence 1, Appl
44	25	4.6	2524	3	US-08-751-965-1	Sequence 1, Appl
45	25	4.6	2524	3	US-08-738-975-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-08-089-998B-1  
; Sequence 1, Application US/0808998B  
; Patent No. 5457178  
; GENERAL INFORMATION:  
; APPLICANT: Jackson et al.  
; TITLE OF INVENTION: INSECTICIDALLY EFFECTIVE SPIDER  
; TITLE OF INVENTION: TOXIN  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz  
; ADDRESSEE: and No. 5457178-1s  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: U.S.A.  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch disk, 720 Kb  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/089,998B  
; FILING DATE: 07-JUL-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: John W. Caldwell  
; REGISTRATION NUMBER: 28,937  
; REFERENCE/DOCKET NUMBER: FMC-0053  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-568-3100  
; TELEFAX: 215-568-3439  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 940 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
US-08-089-998B-1

Query Match 4.8%; Score 26; DB 1; Length 940;  
Best Local Similarity 100.0%; Pred. No. 0.0068;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 517 caatgaaaaa542  
Db 907 CAATGAAAAA932

## RESULT 2

US-08-457-272-1  
; Sequence 1, Application US/08457272  
; Patent No. 5695959  
; GENERAL INFORMATION:  
; APPLICANT: Jackson et al.  
; TITLE OF INVENTION: INSECTICIDALLY EFFECTIVE SPIDER  
; TITLE OF INVENTION: TOXIN  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz  
; ADDRESSEE: and No. 5695959r1s  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: U.S.A.  
; ZIP: 19103

COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch disk, 720 Kb  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/457,272  
; FILING DATE: Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/089,998  
; FILING DATE: 07-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: John W. Caldwell  
; REGISTRATION NUMBER: 28,937  
; REFERENCE/DOCKET NUMBER: FMC-0073  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-568-3100  
; TELEFAX: 215-568-3439  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 940 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; US-08-457-272-1

Query Match 4.8%; Score 26; DB 2; Length 940;  
Best Local Similarity 100.0%; Pred. No. 0.0068;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 517 caatgaaaaa542  
Db 907 CAATGAAAAA932

## RESULT 3

PCT-US94-07595-1  
; Sequence 1, Application PC/TUS9407595  
; GENERAL INFORMATION:  
; APPLICANT: Jackson et al.  
; TITLE OF INVENTION: INSECTICIDALLY EFFECTIVE SPIDER  
; TITLE OF INVENTION: TOXIN  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FMC Corporation  
; STREET: 1735 Market Street  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: U.S.A.

; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch disk, 720 Kb  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/07595  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/089,998  
; FILING DATE: 07 JUL 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Charles C. Fellows &  
; NAME: Robert M. Kennedy  
; REGISTRATION NUMBER: 24,261 & 28,026  
; REFERENCE/DOCKET NUMBER: 4797  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-299-6970  
; TELEFAX: 215-299-6984  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 940 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; PCT-US94-07595-1

Query Match 4.8%; Score 26; DB 6; Length 940;  
Best Local Similarity 100.0%; Pred. No. 0.0068;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 517 caatgaaaaa542  
Db 907 CAATGAAAAA932

## RESULT 4

US-08-900-148-1  
; Sequence 1, Application US/08900148  
; Patent No. 5965792  
; GENERAL INFORMATION:  
; APPLICANT: Schroeder, Julian I.  
; APPLICANT: Antosiewicz, Danuta M.  
; APPLICANT: Schachtman, Daniel P.  
; APPLICANT: Clemens, Stephan  
; TITLE OF INVENTION: Nucleic Acids Encoding Metal Uptake  
; TITLE OF INVENTION: Transporters and Their Uses  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/900,148  
; FILING DATE: 28-JUL-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/022,722  
; FILING DATE: 23-JUL-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774



```
; REFERENCE/DOCKET NUMBER: 023070-087600US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1982 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 137..1471
; OTHER INFORMATION: /product= "low affinity cation
; OTHER INFORMATION: transporter 1 (LCII)"
US-08-900-148-1

Query Match          4.8%; Score 26; DB 4; Length 1982;
Best Local Similarity 100.0%; Pred. No. 0.0067;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 517 caatgaaaaaaatgaaaaaaatgaaaaaa 542
Db 1955 CATTGAAAAAAATGAAAAAAATGAAAAAA 1980

RESULT 5
US-08-924-759-15
; Sequence 15, Application US/08924759
; Patent No. 5962229
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE
; TITLE OF INVENTION: ENZYMES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/924,759
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CL-1128
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 861 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: MAIZE
; IMMEDIATE SOURCE:
```

```
; CLONE: CRIN.PK0014.G8
; US-08-924-759-15

Query Match          4.6%; Score 25; DB 4; Length 861;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 aatgaaaaaaatgaaaaaaatgaaaaaa 542
Db 829 AATGAAAAAAATGAAAAAAATGAAAAAA 853

RESULT 6
US-08-909-965C-9
; Sequence 9, Application US/08909965C
; Patent No. 5936078
; GENERAL INFORMATION:
; APPLICANT: Kuga Tetsuro
; APPLICANT: Nakagawa Satoshi
; APPLICANT: Sakaki yoshiyuki
; APPLICANT: Zhao Nanding
; APPLICANT: Hashida Hideji
; TITLE OF INVENTION: NOVEL DNA, NOVEL POLYPEPTIDE
; TITLE OF INVENTION: AND NOVEL ANTIBODY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FITZPATRICK, CELLA, HARPER AND SCINTO
; STREET: 277 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10172-0194
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/909,965C
; FILING DATE: August 12, 1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 322745/95
; APPLICATION NUMBER: PCT/JP96/03630
; FILING DATE: 12-Dec-1996
; FILING DATE: 12-Dec-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Lawrence S. Peiry
; REGISTRATION NUMBER: 31865
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-2400
; TELEFAX: 212-758-2982
; TELEX: 236262
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 882 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: human
; IMMEDIATE SOURCE:
; CLONE: F1180
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 97 to 195
; IDENTIFICATION METHOD: by experiment
US-08-909-965C-9
```

Query Match 4.6%; Score 25; DB 4; Length 882;

```

Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 aatgaataaataaataaataaataa 542
|||||
DB 826 AATGAAAAAATAAATAAATAA 850

RESULT 7
US-08-839-008-1
; Sequence 1, Application US/08839008
; Patent No. 5916758
; GENERAL INFORMATION:
; APPLICANT: Hurlle, Mark R
; APPLICANT: McDonnell, Peter C
; APPLICANT: McNulty, Dean E
; APPLICANT: Rosen, Craig A
; APPLICANT: Siemens, Ivo R
; APPLICANT: Young, Peter R
; APPLICANT: Yue, Tian-Li
; TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/839,008
; FILING DATE: 23-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/563,697
; FILING DATE: 28-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: P50384
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5096
; TELEFAX: 610-270-5090
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1537 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-839-008-1

Query Match 4.6%; Score 25; DB 4; Length 1537;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 aatgaataaataaataaataaataa 542
|||||
DB 1507 AATGAAAAAATAAATAAATAA 1531

RESULT 8
US-07-841-646-24
; Sequence 24, Application US/07841646
; Patent No. 5266683
; GENERAL INFORMATION:

```

```
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1873 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: MURIDAE
; TISSUE TYPE: EMBRYO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1393
; OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
; OTHER INFORMATION: /product= "MOPI"
; OTHER INFORMATION: /note= "MOPI (CDNA)"
US-07-841-646-24

Query Match 4.6%; Score 25; DB 1; Length 1873;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 518 aatgaaaaaa542
Db 1835 AATGAAAAAA1859

RESULT 9
US-07-901-703-8
; Sequence 8, Application US/07901703
; Patent No. 534654
; GENERAL INFORMATION:
; APPLICANT: RUEGER, DAVID C
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; TITLE OF INVENTION: PROSTHETIC DEVICES HAVING ENHANCED
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: EXCHANGE PLACE, 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,703
; FILING DATE: 19920616
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FITCHER, ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: STK-057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1873 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
```

```
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: MURIDAE
; TISSUE TYPE: EMBRYO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1393
; OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
; OTHER INFORMATION: /product= "MOPI"
; OTHER INFORMATION: /note= "MOPI (CDNA)"
US-07-901-703-8

Query Match 4.6%; Score 25; DB 1; Length 1873;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 518 aatgaaaaaa542
Db 1835 AATGAAAAAA1859

RESULT 10
US-08-147-023-24
; Sequence 24, Application US/08147023
; Patent No. 5468845
; GENERAL INFORMATION:
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: OSTEOGENIC DEVICES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,023
; FILING DATE: 21-FEB-1992
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 810,560
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 827,052
; FILING DATE: 28-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 660,162
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,988
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,849
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 616,374
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 600,024
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 599,543
```



REFERENCE/DOCKET NUMBER: CRP-076FW  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7560  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1873 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 104..1393  
OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"  
OTHER INFORMATION: /product= "mop1-pp"  
OTHER INFORMATION: /note= "mop1 CDNA"  
US-08-480-528A-5

Query Match 4.6%; Score 25; DB 1; Length 1873;  
Best Local Similarity 100.0%; Pred. No. 0.018;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 aatgaaaaa542  
|||||

Db 1835 AATGAAAAA1859

RESULT 13  
US-08-479-666-5  
Sequence 5, Application US/08479666  
Patent No. 5652337  
GENERAL INFORMATION:  
APPLICANT: OPPERMAN, HERMAN  
APPLICANT: OZKAYNAK, ENGIN  
APPLICANT: KUBERASAMPATH, THANGAVEL  
APPLICANT: RUEGER, DAVID C.  
APPLICANT: PANG, ROY H.L.  
APPLICANT: COHEN, CHARLES M.  
TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES  
STREET: 45 SOUTH STREET  
CITY: HOPKINTON  
STATE: MA  
COUNTRY: USA  
ZIP: 01748  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479,666  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FENTON Esq., GILLIAN M.  
REGISTRATION NUMBER: 36,508  
REFERENCE/DOCKET NUMBER: CRP-076DV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (508) 435-9001  
TELEFAX: (508) 435-6951  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1873 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:

NAME/KEY: CDS  
LOCATION: 104..1393  
OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"  
OTHER INFORMATION: /product= "mop1-pp"  
OTHER INFORMATION: /note= "mop1 CDNA"  
US-08-479-666-5

Query Match 4.6%; Score 25; DB 1; Length 1873;  
Best Local Similarity 100.0%; Pred. No. 0.018;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 aatgaaaaa542  
|||||

Db 1835 AATGAAAAA1859

RESULT 14  
US-08-155-343A-18  
Sequence 18, Application US/08155343A  
Patent No. 5656593  
GENERAL INFORMATION:  
APPLICANT: KUBERASAMPATH, THANGAVEL  
APPLICANT: RUEGER, DAVID C.  
APPLICANT: OPPERMAN, HERMAN  
APPLICANT: COHEN, CHARLES M.  
APPLICANT: PANG, ROY H.L.  
TITLE OF INVENTION: MORPHOGENIC-INDUCED PERIODONTAL TISSUE  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES  
STREET: 45 SOUTH STREET  
CITY: HOPKINTON  
STATE: MA  
COUNTRY: USA  
ZIP: 01748  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/155,343A  
FILING DATE: 15-NOV-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FENTON Esq., GILLIAN M.  
REGISTRATION NUMBER: 36,508  
REFERENCE/DOCKET NUMBER: CRP-067FW  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7560  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1873 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:

NAME/KEY: CDS  
LOCATION: 104..1393  
OTHER INFORMATION: /product= "mop1 (CDNA)"  
US-08-155-343A-18

Query Match 4.6%; Score 25; DB 1; Length 1873;  
Best Local Similarity 100.0%; Pred. No. 0.018;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 aatgaaaaa542  
|||||

Search completed: May 1, 2000, 16:04:44  
Job time: 18349 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 19:59:38 ; Search time 4088.29 Seconds  
(without alignments)  
-454.624 Million cell updates/sec

Title: US-09-215-435-78

Perfect score: 542  
Sequence: 1 cagcagctgtggccatgat.....aaaaaaaaaaaaaaaaaaaaa 542

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 5265378 seqs, -1714608768 residues

Word size : 0

Total number of hits satisfying chosen parameters: 10530736

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Pending Patents NA.\*  
1: /cgn2\_6/ptodata/1/pna/PCTUS9\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/US070\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US081A\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US081B\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US081C\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US081D\_COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US082A\_COMB.seq.\*  
9: /cgn2\_6/ptodata/1/pna/US082B\_COMB.seq.\*  
10: /cgn2\_6/ptodata/1/pna/US082C\_COMB.seq.\*  
11: /cgn2\_6/ptodata/1/pna/US083A\_COMB.seq.\*  
12: /cgn2\_6/ptodata/1/pna/US083B\_COMB.seq.\*  
13: /cgn2\_6/ptodata/1/pna/US084A\_COMB.seq.\*  
14: /cgn2\_6/ptodata/1/pna/US084B\_COMB.seq.\*  
15: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*  
16: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*  
17: /cgn2\_6/ptodata/1/pna/US087A\_COMB.seq.\*  
18: /cgn2\_6/ptodata/1/pna/US087B\_COMB.seq.\*  
19: /cgn2\_6/ptodata/1/pna/US087C\_COMB.seq.\*  
20: /cgn2\_6/ptodata/1/pna/US088A\_COMB.seq.\*  
21: /cgn2\_6/ptodata/1/pna/US088B\_COMB.seq.\*  
22: /cgn2\_6/ptodata/1/pna/US088C\_COMB.seq.\*  
23: /cgn2\_6/ptodata/1/pna/US088D\_COMB.seq.\*  
24: /cgn2\_6/ptodata/1/pna/US089A\_COMB.seq.\*  
25: /cgn2\_6/ptodata/1/pna/US089B\_COMB.seq.\*  
26: /cgn2\_6/ptodata/1/pna/US089C\_COMB.seq.\*  
27: /cgn2\_6/ptodata/1/pna/US089D\_COMB.seq.\*  
28: /cgn2\_6/ptodata/1/pna/US089E\_COMB.seq.\*  
29: /cgn2\_6/ptodata/1/pna/US089F\_COMB.seq.\*  
30: /cgn2\_6/ptodata/1/pna/US089G\_COMB.seq.\*  
31: /cgn2\_6/ptodata/1/pna/US089H\_COMB.seq.\*  
32: /cgn2\_6/ptodata/1/pna/US089I\_COMB.seq.\*  
33: /cgn2\_6/ptodata/1/pna/US090A\_COMB.seq.\*  
34: /cgn2\_6/ptodata/1/pna/US090B\_COMB.seq.\*  
35: /cgn2\_6/ptodata/1/pna/US090C\_COMB.seq.\*  
36: /cgn2\_6/ptodata/1/pna/US090D\_COMB.seq.\*  
37: /cgn2\_6/ptodata/1/pna/US090E\_COMB.seq.\*  
38: /cgn2\_6/ptodata/1/pna/US091A\_COMB.seq.\*  
39: /cgn2\_6/ptodata/1/pna/US091B\_COMB.seq.\*  
40: /cgn2\_6/ptodata/1/pna/US091C\_COMB.seq.\*  
41: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq.\*  
42: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq.\*  
43: /cgn2\_6/ptodata/1/pna/US092C\_COMB.seq.\*

44: /cgn2\_6/ptodata/1/pna/US092C\_COMB.seq.\*  
45: /cgn2\_6/ptodata/1/pna/US092D\_COMB.seq.\*  
46: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq.\*  
47: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq.\*  
48: /cgn2\_6/ptodata/1/pna/US093C\_COMB.seq.\*  
49: /cgn2\_6/ptodata/1/pna/US093D\_COMB.seq.\*  
50: /cgn2\_6/ptodata/1/pna/US093E\_COMB.seq.\*  
51: /cgn2\_6/ptodata/1/pna/US093F\_COMB.seq.\*  
52: /cgn2\_6/ptodata/1/pna/US094A\_COMB.seq.\*  
53: /cgn2\_6/ptodata/1/pna/US094B\_COMB.seq.\*  
54: /cgn2\_6/ptodata/1/pna/US094C\_COMB.seq.\*  
55: /cgn2\_6/ptodata/1/pna/US094D\_COMB.seq.\*  
56: /cgn2\_6/ptodata/1/pna/US094E\_COMB.seq.\*  
57: /cgn2\_6/ptodata/1/pna/US094F\_COMB.seq.\*  
58: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq.\*  
59: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq.\*  
60: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq.\*  
61: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq.\*  
62: /cgn2\_6/ptodata/1/pna/US095E\_COMB.seq.\*  
63: /cgn2\_6/ptodata/1/pna/US095F\_COMB.seq.\*  
64: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq.\*  
65: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq.\*  
66: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq.\*  
67: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq.\*  
68: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq.\*  
69: /cgn2\_6/ptodata/1/pna/US096F\_COMB.seq.\*  
70: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq.\*  
71: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq.\*  
72: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq.\*  
73: /cgn2\_6/ptodata/1/pna/US097D\_COMB.seq.\*  
74: /cgn2\_6/ptodata/1/pna/US097E\_COMB.seq.\*  
75: /cgn2\_6/ptodata/1/pna/US097F\_COMB.seq.\*  
76: /cgn2\_6/ptodata/1/pna/US098A\_COMB.seq.\*  
77: /cgn2\_6/ptodata/1/pna/US098B\_COMB.seq.\*  
78: /cgn2\_6/ptodata/1/pna/US098C\_COMB.seq.\*  
79: /cgn2\_6/ptodata/1/pna/US098D\_COMB.seq.\*  
80: /cgn2\_6/ptodata/1/pna/US098E\_COMB.seq.\*  
81: /cgn2\_6/ptodata/1/pna/US098F\_COMB.seq.\*  
82: /cgn2\_6/ptodata/1/pna/US098G\_COMB.seq.\*  
83: /cgn2\_6/ptodata/1/pna/US098H\_COMB.seq.\*  
84: /cgn2\_6/ptodata/1/pna/US098I\_COMB.seq.\*  
85: /cgn2\_6/ptodata/1/pna/US099A\_COMB.seq.\*  
86: /cgn2\_6/ptodata/1/pna/US099B\_COMB.seq.\*  
87: /cgn2\_6/ptodata/1/pna/US099C\_COMB.seq.\*  
88: /cgn2\_6/ptodata/1/pna/US099D\_COMB.seq.\*  
89: /cgn2\_6/ptodata/1/pna/US099E\_COMB.seq.\*  
90: /cgn2\_6/ptodata/1/pna/US099F\_COMB.seq.\*  
91: /cgn2\_6/ptodata/1/pna/US099G\_COMB.seq.\*  
92: /cgn2\_6/ptodata/1/pna/US099H\_COMB.seq.\*  
93: /cgn2\_6/ptodata/1/pna/US099I\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	542	100.0	542	42	US-09-215-435-78
2	542	100.0	565	73	US-60-096-116-61
3	429	79.2	523	54	PCT-US99-443-634-1
4	298	55.0	680	1	US-09-489-847-73
5	298	55.0	680	92	US-09-489-847-73
6	287	53.0	433	93	US-60-184-698-48
7	287	53.0	478	93	US-60-184-698-694
8	228	42.1	421	25	US-08-905-279-44
9	202	37.3	488	7	US-08-196-481-520
10	202	37.3	488	7	US-08-196-481-520
11	176	32.5	263	33	US-08-992-332-1466
12	176	32.5	263	61	US-60-033-364-14
13	147	27.1	514	84	US-60-147-499-1374

Sequence 78, Appl  
Sequence 61, Appl  
Sequence 1, Appl  
Sequence 73, Appl  
Sequence 48, Appl  
Sequence 694, Appl  
Sequence 44, Appl  
Sequence 520, Appl  
Sequence 1466, Appl  
Sequence 14, Appl  
Sequence 1374, Appl





```
; APPLICATION NUMBER: US/60/096.116
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.038PR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Testis
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 16..84
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 9.8
; OTHER INFORMATION: seq FLLFFFLFLTRG/SL
; FEATURE:
; NAME/KEY: poly_a_signal
; LOCATION: 502..507
; FEATURE:
; NAME/KEY: poly_a
; LOCATION: 529..542
; IDENTIFICATION METHOD: blastn2
; FEATURE:
; NAME/KEY: vrt
; LOCATION: complement(33..74)
; IDENTIFICATION METHOD: fasta
; OTHER INFORMATION: Identity 88
; OTHER INFORMATION: region 21722..21764
; OTHER INFORMATION: id AF048728
; FEATURE:
; NAME/KEY: vrt
; LOCATION: 511..545
; IDENTIFICATION METHOD: fasta
; OTHER INFORMATION: Identity 83
; OTHER INFORMATION: region 4627..4663
; OTHER INFORMATION: id AF048728
US-60-096-116-61
```

```
Query Match 100.0%; Score 542; DB 73; Length 565;
Best Local Similarity 100.0%; Pred. No. 7.6e-217;
Matches 542; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cagcaccctgtggccatgatctaccacaatgctgctgctgctgcttctctcttcttc 60
Db 1 CAGGACCTGTGGCCCATGATGCTACCCCAATGGCTGCTGCTGCTGCTGCTGCTGCT 60

Qy 61 ttctctctctctctcaccagggtctcactttctcccaaaaaataaaaccttttgagctc 120
Db 61 TTCTCTTCTCTCTCTCCTCACCAGGGGCTCACCTTCTCCAAATAAATACCTTTTGAGCTC 120

Qy 121 aaggagcttgatcccggaacagagactcgagactggtgctgctgcaacgtctccagac 180
Db 121 AAGGAGCTCTTGATCCCGGAACAGAGACTCGAGACTGGCTGCTGCCAAGCTCTCCAGAC 180

Qy 181 aattcgagctcgactcgccgagaggggtccgagggcagctctgtgtcaaacgaggtg 240
Db 181 AATTGCGAGCTCGACTGCGCGGAGAGGGGTCCGAGGGGAGTCTGTGTCAAAACGCGGTG 240

Qy 241 ttctttggccaatatagagcgtgtccctgcctgcggaaacctgactgttatattcaag 300
Db 241 TTCTTTGGCCATATAGAGCGTGTCCCTGCCCTGGGAACCTGACTGTATATATTCAAAG 300
```

```
Qy 301 aatgagaaatggcttagcatcgccctatggccgttgttcagaaaaattggagcagaagttg 360
Db 301 AATGAGAAATGGCTTAGCATCGCCCTATGGCCGTGTGTCAGAAAAATGGAAGCGAGAAGTTG 360

Qy 361 gctaagaaaatgttcttcttagtgcctcctcttcttctgctcctcctcctcctcctgc 420
Db 361 GCTAAGAAAATGTTCTTCTTAGTGCCTCCTCCTCTTGTGTGCTCTCTCTCTCCACCTGC 420

Qy 421 tctctcctaccagagactgtgtttcacccctgttccccagagctccaccatgagtgg 480
Db 421 TCTCTCCTCCTACCAGAGACTGTGTGTTTCACTCTGTTCCTCCAGAGCTCCACCATGAGTGG 480

Qy 481 agggaaagtgggagtgattgaaataaagagctttttcaatgaaaaaataaaaaa 540
Db 481 AGGGAAGTGGGAGTGATTGAAATAAGAGCTTTTCAATGAAAAAATAAAAAA 540

Qy 541 aa 542
Db 541 AA 542
```

```
RESULT 3
US-09-443-634-1
; Sequence 1, Application US/09443634
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: Colipase-Like Protein - 1
; FILE REFERENCE: 99-85
; CURRENT APPLICATION NUMBER: US/09/443,634
; CURRENT FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 523
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(354)
US-09-443-634-1

Query Match 79.2%; Score 429; DB 54; Length 523;
Best Local Similarity 100.0%; Pred. No. 1.4e-169;
Matches 429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 114 ggagctcaagagctcttgcctccggaacccagagactgcgagactggtgctgccaacgtgc 173
Db 90 ggagctcaagagctcttgcctccggaacccagagactgcgagactggtgctgccaacgtgc 149

Qy 174 tccagacaattgcgagtcgactgcgagagaagggtccgagggcagctctgtgtcaaac 233
Db 150 tccagacaattgcgagtcgactgcgagagaagggtccgagggcagctctgtgtcaaac 209

Qy 234 gcaggtgtctttggccaatatagagctgtccctgcctgcctgcggaaacctgactgtata 293
Db 210 gcaggtgtctttggccaatatagagctgtccctgcctgcctgcggaaacctgactgtata 269

Qy 294 ttcaagaatgagaaaatgcttagcatcgcttatgcccgttgtcagaaaaattggaagcca 353
Db 270 ttcaagaatgagaaaatgcttagcatcgcttatgcccgttgtcagaaaaattggaagcca 329

Qy 354 gaagtgtgcttaagaaaatgttcttcttagtgcctcctctcttcttgcctcctcctcctc 413
Db 330 gaagtgtgcttaagaaaatgttcttcttagtgcctcctctcttcttgcctcctcctcctc 389

Qy 414 cactgtctctcctcctaccagagctgtgtttcacccctgttccccagagcctccacca 473
Db 390 cactgtctctcctcctaccagagctgtgtttcacccctgttccccagagcctccacca 449

Qy 474 tgagtggaggaagtgaggagtgattgaataaagagcttttcaatgaaaaaataaaaa 533
```

```
Db 450 tgagtgagggaggtgggaggtgattgaaataaagagctttttcaatgnaaaaaaaaaa 509
QY 534 aaaaaaaa 542
Db 510 aaaaaaaa 518

RESULT 4
PCT-US99-17130-73
; Sequence 73, Application PC/TUS9917130
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: P2031.PCT
; CURRENT APPLICATION NUMBER: PCT/US99/17130
; CURRENT FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 73
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (9)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (15)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (16)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (22)
; OTHER INFORMATION: n equals a,t,g, or c
PCT-US99-17130-73

Query Match 55.0%; Score 298; DB 1; Length 680;
Best Local Similarity 99.5%; Pred. No. 9e-115;
Matches 398; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 cagcactgtggggccatgatgtaccaccaatggctgctgctgtgttctctcttc 60
Db 118 cagcactgtggggccatgatgtaccaccaatggctgctgctgtgttctctcttc 177

QY 61 ttcttctctctaccagggtgcactttctccaaaaataaaccttttggagctc 120
Db 178 ttcttctctctaccagggtgcactttctccaaaaataaaccttttggagctc 237

QY 121 aaggagcttgatccggaaaccagagactgcgagactgctgctgctgctgctcagac 180
Db 238 aaggagcttgatccggaaaccagagactgcgagactgctgctgctgctgctcagac 297

QY 181 aattcgagtcgactgcgcgggagagggtcccgaggcagctgtgtgtcaaacgcaggtg 240
```

```
Db 298 aattcgagtcgactgcgcgggagagggtcccgagggtcgagtcgtgtgtcaaacgcaggtg 357
QY 241 ttcttggccaaatagagcgtgtccctgcctgcctgcggaacctgactgtgtatattcaag 300
Db 358 ttcttggccaaatagagcgtgtccctgcctgcctgcggaacctgactgtgtatattcaag 417
QY 301 aatgagaaatggcttagcatcgcctatgctgctgttcagaaaaattgggaaggcagaagtgtg 360
Db 418 aatgagaaatggcttagcatcgcctatgctgctgttcagaaaaattgggaaggcagaagtgtg 477
QY 361 gctaagaaaaatgtttcttcttagtgcctccctctctctctctctctctctctctctctctg 400
Db 478 gctaagaaaaatgtttcttcttagtgcctccctctctctctctctctctctctctctctg 517

RESULT 5
US-09-489-847-73
; Sequence 73, Application US/09489847
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: P2031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 73
; LENGTH: 680
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (9)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (15)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (16)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (22)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-489-847-73
```

```
Query Match 55.0%; Score 298; DB 92; Length 680;
Best Local Similarity 99.5%; Pred. No. 9e-115;
Matches 398; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 cagcactgtggggccatgatgtaccaccaatggctgctgctgtgttctctcttc 60
Db 118 cagcactgtggggccatgatgtaccaccaatggctgctgctgtgttctctcttc 177
```



```

Query Match      42.1%; Score 228; DB 25; Length 421;
Best Local Similarity 100.0%; Pred. NO. 1.8e-85;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 ggaacacaggactgcgagaactggctgctgcgaacgtgtccagacaaatgcgagtgccact 196
      |||||
Db 183 GGAACcAGGACTGGAGAGTGGCTGCTGCCAACGTGTCCAGACAATTCGAGTGCcACT 242
      |||||

QY 197 qcgcgagaaagggtccgagggcagtgctgtcaaacccaggtgtctcttggccaata 256
      |||||
Db 243 GCGCGGAGAAAGGGTCCGAGGGCAGTGTGTCAAAACCAGGTGTCTTTGGCCAAATA 302

QY 257 ggcggtgccctgcctgcggaaactgactgtgtatatattcaagaatgagaaatggccta 316
      |||||
Db 303 GAGCGTGTCCCTGCTGCGGGAACCTGACTGTGTATATTATCAAGAATGAGAAATGGCTTA 362

QY 317 gcacgcgcctatggccgttgcagaaaaattggaaggcagaagtggcta 364
      |||||
Db 363 GCATCGCCTATGGCGGTTGTGAGAAAAATTGGAAGGCAGAAGTTGGCTA 410

RESULT 9

```



```
QY 174 tcacagcaattgcagctgcactgcgcgagagaggggtccgagggcagctgtgtgtcaaac 233
Db 120 TCACACAAATGCGAGTCCGACTGCGCGAGAGGGGTCCGAGGGCAGTCTGTGTCAAAAC 179

QY 234 gcaggtgtttttggccaatatagacgtgtccctgcctgcgcggaacctgactgtga 289
Db 180 GCAGGTGTTCTTTGGCCAATATAGAGCGTGTCCCTGCTGCGGAACCTGACTGTGTA 235

RESULT 12
US-60-033-364-14
; Sequence 14, Application US/60033364
; GENERAL INFORMATION:
; APPLICANT: Gooding, Douglas H.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Akerblom, Ingrid E.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Naughton, Rebecca E.
; APPLICANT: Klingler, Tod M.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM
; TITLE OF INVENTION: HUMAN PITUITARY
; NUMBER OF SEQUENCES: 1416
; CORRESPONDENCE ADDRESS:
; ADDRESS: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/033,364
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PD-0291P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 855-0555
; TELEFAX: (415) 845-4166
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: 2084384H1
; US-60-033-364-14

Query Match 32.5%; Score 176; DB 61; Length 263;
Best Local Similarity 100.0%; Pred. No. 1e-63;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ggagctcaagagtgcttgcataccggaaccaggactgcgagactggctgctgccaaactgc 173
Db 60 GGAGCTCAAGAGTCTTGCATCCGGAACAGGACTGCGAGACTGCGTCTGCCAACGTGC 119

QY 174 tcacagcaattgcagctgcactgcgcgagaggggtccgagggcagctgtgtgtcaaac 233
Db 120 TCACACAAATGCGAGTCCGACTGCGCGAGAGGGGTCCGAGGGCAGTCTGTGTCAAAAC 179

QY 234 gcaggtgtttttggccaatatagacgtgtccctgcctgcgcggaacctgactgtga 289
Db 180 GCAGGTGTTCTTTGGCCAATATAGAGCGTGTCCCTGCTGCGGAACCTGACTGTGTA 235

Query Match 27.1%; Score 147; DB 84; Length 514;
Best Local Similarity 99.5%; Pred. No. 1.4e-51;
Matches 197; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 cagcacctgtgggcccattgctaccaccaatggctcaccatggctgctgctgtctctctcttc 60
Db 84 cagcacctgtgggcccattgctaccaccaatggctgctgctgctgtctctctctctcttc 143

QY 61 ttctctctctctcaccagggggtcactttctccacaaaataataacacacacacacacacac 120
Db 144 ttctctctctcaccagggggtcactttctccacaaaataataacacacacacacacacac 203

QY 121 aaggagctgttgcataccggaaccaggactgcgagactggctgctgctgctgctgctgctg 180
Db 204 aaggagctgttgcataccggaaccaggactgcgagactggctgctgctgctgctgctgctg 263

QY 181 aattgcagctgcactgc 198
Db 264 aattgcagctgcactgc 281

RESULT 14
US-60-169-629-108
; Sequence 108, Application US/60169629
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Bougueleret, L.
; APPLICANT: Jobert, S.
; TITLE OF INVENTION: cDNAs for Secreted Proteins
; FILE REFERENCE: GENSET.071PRF
; CURRENT APPLICATION NUMBER: US/60/169,629
; CURRENT FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 715
; SOFTWARE: Patent.pm
; SEQ ID NO 108
; LENGTH: 514
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 113..364
```

;; FEATURE: sig\_peptide  
;; NAME/KEY: 113..172  
;; LOCATION: Von Heijne matrix  
;; OTHER INFORMATION: score 4.3  
;; OTHER INFORMATION: seq SLLSLPPHQGLT/FS  
;; FEATURE: polyA\_signal  
;; NAME/KEY: 478..483  
;; LOCATION: 478..483  
;; FEATURE: polyA\_site  
;; NAME/KEY: 500..514  
;; LOCATION: 500..514  
US-60-169-629-108

Query Match 27.1%; Score 147; DB 86; Length 514;  
Best Local Similarity 99.5%; Pred. No. 1.4e-51;  
Matches 197; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 cagcaccctgtggccatgatgtaccccaatggctgctgctgttctcttcttcttc 60  
Db 84 cagcaccctgtggccatgatgtaccccaatggctgctgctgttctcttcttcttc 143  
Qy 61 ttctctctctctcaccagggtcactttctccaaataataaacttttggagctc 120  
Db 144 ttctctctctctcaccagggtcactttctccaaataataaacttttggagctc 203  
Qy 121 aaggagtcttgcacccgagcagcagcagcagcagcagcagcagcagcagcagc 180  
Db 204 aaggagtcttgcacccgagcagcagcagcagcagcagcagcagcagcagcagc 263  
Qy 181 aattgcgagtcgactgc 198  
Db 264 aattgcgagtcgactgc 281

RESULT 15  
US-60-187-470-108  
; Sequence 108, Application US/60187470  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Bougueleret, L.  
; APPLICANT: Jobert, S.  
; TITLE OF INVENTION: cDNAs for Secreted Proteins  
; FILE REFERENCE: 78.US2.PRO  
; CURRENT APPLICATION NUMBER: US/60/187.470  
; CURRENT FILING DATE: 2000-03-06  
; NUMBER OF SEQ ID NOS: 715  
; SOFTWARE: Patent.pm  
; SEQ ID NO 108  
; LENGTH: 514  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 113..364  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 113..172  
; OTHER INFORMATION: Von Heijne matrix  
; OTHER INFORMATION: score 4.3  
; OTHER INFORMATION: seq SLLSLPPHQGLT/FS  
; FEATURE:  
; NAME/KEY: polyA\_signal  
; LOCATION: 478..483  
; FEATURE:  
; NAME/KEY: polyA\_site  
; LOCATION: 500..514  
US-60-187-470-108

Query Match 27.1%; Score 147; DB 93; Length 514;  
Best Local Similarity 99.5%; Pred. No. 1.4e-51;

Matches 197; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 cagcaccctgtggccatgatgtaccccaatggctgctgctgttctcttcttcttc 60  
Db 84 cagcaccctgtggccatgatgtaccccaatggctgctgctgttctcttcttcttc 143  
Qy 61 ttctctctctctcaccagggtcactttctccaaataataaacttttggagctc 120  
Db 144 ttctctctctctcaccagggtcactttctccaaataataaacttttggagctc 203  
Qy 121 aaggagtcttgcacccgagcagcagcagcagcagcagcagcagcagcagcagc 180  
Db 204 aaggagtcttgcacccgagcagcagcagcagcagcagcagcagcagcagcagc 263  
Qy 181 aattgcgagtcgactgc 198  
Db 264 aattgcgagtcgactgc 281

Search completed: May 1, 2000, 19:59:41  
Job time: 21965 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 14:44:22 ; Search time 3022.95 Seconds  
(without alignments)  
676.958 Million cell updates/sec

Title: US-09-215-435-78  
Perfect score: 542  
Sequence: 1 cagcactgtggccatgat.....aaaaaaaaaaaaaaaaaaaa 542

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4538634 seqs, 1887831982 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : EST:  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*

44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_gss1:\*  
80: gb\_gss2:\*  
81: gb\_gss3:\*  
82: gb\_gss4:\*  
83: em\_gss1:\*  
84: em\_gss2:\*  
85: em\_gss3:\*  
86: em\_gss4:\*  
87: gb\_gss5:\*  
88: gb\_gss6:\*  
89: gb\_gss7:\*  
90: gb\_gss8:\*  
91: gb\_gss9:\*  
92: em\_gss5:\*  
93: em\_gss6:\*  
94: em\_gss7:\*  
95: em\_gss8:\*  
96: em\_gss9:\*  
97: em\_gss10:\*  
98: em\_gss11:\*  
99: gb\_gss10:\*  
100: gb\_gss11:\*  
101: em\_gss12:\*  
102: gb\_gss12:\*  
103: gb\_gss13:\*  
104: gb\_gss14:\*  
105: gb\_gss15:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match	Length	DB	ID	Description
No.	Score					
c 1	199	36.7	491	41	AI018488	AI018488 ov63f08.s



Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/dbp/image/image.html](http://www-bio.llnl.gov/dbp/image/image.html)

Insert Length: 963 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 401.

#### FEATURES

source  
 1. .875  
 /location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1284555"  
 /clone\_lib="NCI-CGAP\_GCB1"  
 /lab\_host="DH10B"  
 /tissue\_type="germinal center B cell"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer  
 [5'-TGTTACCAATCTGAAGTGGCGGCGCTCATTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."  
 249 a 193 c 148 g 285 t

#### BASE COUNT

ORIGIN  
 249 a 193 c 148 g 285 t

Query Match 5.9%; Score 32; DB 38; Length 875;  
 Best Local Similarity 100.0%; Pred. No. 0.00089;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 cttttcaatgaaaaaa 542  
 |||||  
 Db 44 CTTTTCATGAAAAA 13

#### RESULT 3

LOCUS A1426351 562 bp mRNA EST 09-MAR-1999  
 DEFINITION mg95e05.x1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA clone IMAGE:437888 3' similar to gb:U08215 Mus musculus Hsp70-related NST-1 (MOUSE);, mRNA sequence.  
 ACCESSION A1426351  
 VERSION A1426351.1 GI:4272282  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 562)  
 Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
 The WashU-NCI Mouse EST Project 1999  
 Unpublished (1999)  
 On May 7, 1998 this sequence version replaced gi:3119377.  
 Contact: Marra M/WashU-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800

Fax: 314 286 1810  
 Email: mouseest@wustl.wustl.edu  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](http://info.image.llnl.gov)) for further information.  
 This clone was previously sequenced on the 5' end only, this new data is from the 3' end  
 High quality sequence stop: 395.

#### FEATURES

Location/Qualifiers  
 source  
 1. .562  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:437888"  
 /clone\_lib="Soares mouse embryo NbME13.5 14.5"  
 /sex="unknown"  
 /tissue\_type="embryo"  
 /dev\_stage="13.5-14.5dpc total fetus"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGCGGCGGAAATTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 ]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."  
 143 a 137 c 109 g 173 t

#### BASE COUNT

ORIGIN  
 143 a 137 c 109 g 173 t

Query Match 5.7%; Score 31; DB 46; Length 562;  
 Best Local Similarity 100.0%; Pred. No. 0.0035;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 cttttcaatgaaaaaa 541  
 |||||  
 Db 31 CTTTTCATGAAAAA 1

#### RESULT 4

LOCUS A1146643 454 bp mRNA EST 29-SEP-1998  
 DEFINITION or32e03.x1 NCI-CGAP\_GC3 Homo sapiens cDNA clone IMAGE:1597564 3', mRNA sequence.  
 ACCESSION A1146643  
 VERSION A1146643.1 GI:3674325  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 454)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 Unpublished (1997)  
 On Jan 19, 1998 this sequence version replaced gi:2045069.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/dbp/image/image.html](http://www-bio.llnl.gov/dbp/image/image.html)

Seq primer: -40ml3 fwd. ET from Amersham.

# FEATURES

Location/Qualifiers  
1. .454  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1597564"  
/clone\_lib="NCI\_CGAP\_GC3"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from 3 pooled germ cell tumors, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is not normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 146 a 86 c 119 g 103 t  
ORIGIN

Query Match 5.5%; Score 30; DB 42; Length 454;  
Best Local Similarity 100.0%; Pred. No. 0.0068;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 513 ttctcaatgaaaaaataaaaaaataaaaaa 542  
|||||  
Db 34 TTTCAATGAAAAAATAAAAAAATAAAAAA 5

## RESULT 5

AW156348  
LOCUS AW156348 346 bp mRNA EST 04-NOV-1999  
DEFINITION se22h08.y1 Gm-cl015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
Gm-cl015-2008 5', mRNA sequence.  
ACCESSION AW156348  
VERSION AW156348.1 GI:6227749  
KEYWORDS EST.  
SOURCE soybean.  
ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

REFERENCE 1 (bases 1 to 346)  
AUTHORS Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.  
TITLE Public Soybean EST Project  
JOURNAL Unpublished (1999)  
COMMENT On Dec 20, 1995 this sequence version replaced gi:1134090.  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 421-3222 FAX: (888) 919-3324 or (314) 421-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com  
Seq primer: -40RP from Gibco  
High quality sequence stop: 304.  
Location/Qualifiers  
1. .346  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-cl015-2008"

## FEATURES

source  
Location/Qualifiers  
1. .542  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2338125"  
/clone\_lib="NCI\_CGAP\_Lu24"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI\_CGAP\_Lu5 was prepared, and ss circles were

/clone\_lib="Gm-cl015"  
/tissue\_type="Mature flowers, field grown plants"  
/lab\_host="XL10-Gold"  
/note="Vector: pBluescript II XR; Site\_1: EcoRI; Site\_2: XhoI; This cDNA library was constructed from mRNA isolated from mature flowers of field grown plants. The cDNA library was prepared using the Strategene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adaptors were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."

BASE COUNT 167 a 37 c 43 g 99 t  
ORIGIN

Query Match 5.4%; Score 29; DB 70; Length 346;  
Best Local Similarity 100.0%; Pred. No. 0.019;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 514 ttctcaatgaaaaaataaaaaaataaaaaa 542  
|||||  
Db 258 TTTCAATGAAAAAATAAAAAAATAAAAAA 286

## RESULT 6

AW1703304/c  
LOCUS AW1703304 542 bp mRNA EST 03-JUN-1999  
DEFINITION w882f11.x1 NCI\_CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:2338125 3',  
mRNA sequence.  
ACCESSION AW1703304  
VERSION AW1703304.1 GI:4991204  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 542)  
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187898.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco  
High quality sequence stop: 457.

## FEATURES

source  
Location/Qualifiers  
1. .542  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2338125"  
/clone\_lib="NCI\_CGAP\_Lu24"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI\_CGAP\_Lu5 was prepared, and ss circles were

made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 206 a 74 c 75 g 187 t  
ORIGIN

Query Match 5.4%; Score 29; DB 50; Length 542;  
Best Local Similarity 100.0%; Pred. No. 0.018;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 514 ttccaatgaaaaaa542

Db 30 TTTCAATGAAAAA542

RESULT 7

AQ538183/c

LOCUS

DEFINITION AQ538183 712 bp DNA GSS 18-MAY-1999

ACCESSION AQ538183

VERSION AQ538183.1 GI:4849873

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

AUTHORS Zhao S., Adams M.D., Nierman W., Malek J., de Jong P. and

Venter J.C.

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

JOURNAL Map Building

COMMENT Unpublished (1997)

Other\_GSSs: RPCI-11-356H4.TJ

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbeetgr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (Info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html.

Seq primer: 17

Class: BAC ends.

FEATURES Location/Qualifiers

1..712

/organism="Homo sapiens"

/db\_xref="GDB:7636491"

/db\_xref="taxon:9606"

/clone="RPCI-11-356H4"

/clone\_lib="RPCI-11"

/sex="Male"

/cell\_type="Lymphocytes"

/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;

RPCI11 Human Male BAC Library"

BASE COUNT 194 a 148 c 137 g 233 t

ORIGIN

Query Match 5.4%; Score 29; DB 104; Length 712;  
Best Local Similarity 100.0%; Pred. No. 0.017;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 514 ttccaatgaaaaaa542

Db 104 TTTCAATGAAAAA76

RESULT 8

AW151320/c

LOCUS

DEFINITION AW151320 172 bp mRNA EST 03-NOV-1999

ACCESSION AW151320

VERSION AW151320.1 GI:6199218

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

AUTHORS 1 (bases 1 to 172)

TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

Unpublished (1997)

On Jun 5, 1998 this sequence version replaced gi:3188335.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: polyt not found

Seq primer: -40UP from Gibco

High quality sequence stop: 157.

FEATURES source

1..172

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2630753"

/clone\_lib="NCI-CGAP\_Utl1"

/tissue\_type="well-differentiated endometrial

adenocarcinoma, 7 pooled tumors"

/lab\_host="DH10B"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI;

Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.75 kb. Life Technologies catalog #:

11538-014"

BASE COUNT 38 a 29 c 47 g 58 t

ORIGIN

Query Match 5.2%; Score 28; DB 70; Length 172;

Best Local Similarity 100.0%; Pred. No. 0.054;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 515 ttccaatgaaaaaa542

Db 29 TTTCAATGAAAAA542

RESULT 9

AI345689

LOCUS

DEFINITION AI345689 196 bp mRNA EST 08-APR-1999

ACCESSION AI345689

VERSION AI345689.1 GI:4082895

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 196)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced gi:2282164.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
cDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Insert Length: 254 Std Error: 0.00  
Seq primer: -40UP from Gibco.

## FEATURES

source  
1. .196  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2061008"  
/clone\_lib="NCI\_CGAP\_Lu26"  
/tissue\_type="invasive adenocarcinoma"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pAMP1; mRNA made from lung  
adenocarcinoma tissue, cDNA made by oligo-dT priming.  
Directionally cloned. Size-selected on agarose gel,  
average insert size 500 bp. Primary library,  
non-amplified."

BASE COUNT 88 a 45 c 38 g 25 t  
ORIGIN

Query Match 5.2%; Score 28; DB 45; Length 196;  
Best Local Similarity 100.0%; Pred. No. 0.053;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 515 ttcaatgaaaaaataaaaaaataaaaaa 542  
|||||

Db 148 TTCAATGAAAAAATAAAAAAATAAAAAA 175

RESULT 10  
AI824533/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE

AI824533 228 bp mRNA EST 24-AUG-1999  
tx71605.x1 NCI\_CGAP\_Ut1 Homo sapiens cDNA clone IMAGE:2275040 3',  
mRNA sequence.  
AI824533  
AI824533.1 GI:5445204  
EST.  
human.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 228)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On May 18, 1998 this sequence version replaced gi:3136716.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Seq primer: -40UP from Gibco  
High quality sequence stop: 170.

## FEATURES

source  
1. .228  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2275040"  
/clone\_lib="NCI\_CGAP\_Ut1"  
/tissue\_type="well-differentiated endometrial  
adenocarcinoma, 7 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.75 kb. Life Technologies catalog #:  
11538-014"

BASE COUNT 51 a 43 c 54 g 80 t  
ORIGIN

Query Match 5.2%; Score 28; DB 61; Length 228;  
Best Local Similarity 100.0%; Pred. No. 0.052;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 515 ttcaatgaaaaaataaaaaaataaaaaa 542  
|||||

Db 38 TTCAATGAAAAAATAAAAAAATAAAAAA 11

RESULT 11  
AW073660  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AW073660 340 bp mRNA EST 13-OCT-1999  
xa58h10.x1 NCI\_CGAP\_HSC2 Homo sapiens cDNA clone IMAGE:2571043 3',  
mRNA sequence.  
AW073660  
AW073660.1 GI:6028658  
EST.  
human.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 340)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Feb 18, 1999 this sequence version replaced gi:4297602.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Herbert Morse, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Possible reversed clone: polyT not found  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1. .340  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2571043"  
/clone\_lib="NCI\_CGAP\_HSC2"  
/tissue\_type="stem cell 34+/38+"  
/dev\_stage="adult"  
/lab\_host="DH10B"

/note="Organ: bone marrow; Vector: pAMP1; mRNA made from bone marrow, stem cells 34+/38", cDNA made by oligo-dT priming. Directionally cloned. Size-selected on agarose gel, average insert size 400 bp. Primary library, non-amplified."

BASE COUNT 118 a 52 c 53 g 117 t  
ORIGIN

Query Match 5.2%; Score 28; DB 64; Length 340;  
Best Local Similarity 100.0%; Pred. No. 0.049;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 515 tcaatgaaaaa542  
Db 296 TTCAATGAAAAA323

## RESULT 12

T65117/c 90 bp mRNA EST 07-MAR-1995  
LOCUS yC74b11.s1 Soares infant brain 1N1B Homo sapiens cDNA clone  
DEFINITION IMAGE:21556 3', mRNA sequence.

ACCESSION T65117 GI:674162  
VERSION T65117.1  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 90)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevisan, E., Waterston, R., Williamson, A., Wohldmann, P. and  
Wilson, R.

## TITLE

The WashU-Werck EST Project  
Unpublished (1995)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu

## Insert Size: 1525

High quality sequence stops: 76 Source: IMAGE Consortium, LLNL This  
clone is available royalty-free through LLNL; contact the IMAGE  
Consortium (info@image.llnl.gov) for further information.

Insert Length: 1525 Std Error: 0.00

Seq primer: -21m13

High quality sequence stop: 76.

## FEATURES

## source

1..90  
/organism="Homo sapiens"  
/db\_xref="GDB:39303"  
/db\_xref="taxon:9606"  
/clone="IMAGE:21556"  
/clone\_lib="Soares infant brain 1N1B"  
/sex="female"  
/dev\_stage="73 days post natal"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: whole brain; Vector: lambda BA; Site\_1: Not  
I; Site\_2: Hind III; 1st strand cDNA was primed with a Not  
I - oligo(dT) primer [5];  
AACTGGAGAAATCGCGCGGAGGATTTTTTTTTTTT 3';  
double-stranded cDNA was ligated to Hind III adaptors  
(Pharmacia), digested with Not I and directionally cloned  
into the Not I and Hind III sites of the lambda BA vector.  
Library went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

## BASE COUNT

## ORIGIN

30 a 10 c 8 g 42 t

Query Match 5.0%; Score 27; DB 21; Length 90;  
Best Local Similarity 100.0%; Pred. No. 0.16;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 tcaatgaaaaa542

Db 31 TCATGAAAAA5

## RESULT 13

AA785723 169 bp mRNA EST 31-JUL-1998  
LOCUS h1f04a1.r1 Aspergillus nidulans 24hr asexual developmental and  
DEFINITION vegetative cDNA lambda zap library Emericella nidulans cDNA clone  
h1f04a1 5', mRNA sequence.

ACCESSION AA785723 GI:2845891

VERSION AA785723.1

KEYWORDS EST.

SOURCE Emericella nidulans.

ORGANISM Emericella nidulans

Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae;

Emericella.

REFERENCE 1 (bases 1 to 169)

AUTHORS Kupfer, D., Gray, J., Hausner, J., Lai, H., Martin, W., Aramayo, R.,

Prade, R. and Roe, B.

TITLE An Aspergillus nidulans EST Database

JOURNAL Unpublished (1998)

COMMENT On Jan 17, 1998 this sequence version replaced gi:2043274.

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu

Department of Chemistry and Biochemistry

Advanced Center for Genome Technology, University of Oklahoma

620 Parrington Oval, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal

Genetics Stock Center

Seq primer: SK

High quality sequence stop: 141.

FEATURES

## source

1..169  
/organism="Emericella nidulans"  
/strain="FGSC A26"  
/db\_xref="taxon:5072"  
/clone="h1f04a1"  
/clone\_lib="Aspergillus nidulans 24hr asexual  
developmental and vegetative cDNA lambda zap library"  
/tissue\_type="vegetative mycelia, asexual structures"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript  
3' end of cDNA cloned into XhoI site of pBluescript"

BASE COUNT 72 a 31 c 22 g 44 t

## ORIGIN

Query Match 5.0%; Score 27; DB 38; Length 169;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 tcaatgaaaaa542

Db 127 TCATGAAAAA153

## RESULT 14

AI343176 210 bp mRNA EST 16-FEB-1999  
LOCUS tb04h10.x2 NC1\_CGAP\_Lu26 Homo sapiens cDNA clone IMAGE:2052643 3',  
DEFINITION mRNA sequence.

ACCESSION AI343176

VERSION AI343176.1

KEYWORDS EST.

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 210)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 17, 1998 this sequence version replaced gi:2044053.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 CDNA Library Preparation: David B. Krizman, Ph.D.  
 CDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 319 Std Error: 0.00

Seq primer: -40UP from Gibco.

Location/Qualifiers

1. .210  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:2052643"  
 /clone\_lib="NCI-CGAP\_Lu26"  
 /tissue\_type="invasive adenocarcinoma"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"

/note="Organ: lung; Vector: pAMP1; mRNA made from lung  
 adenocarcinoma tissue, cDNA made by oligo-dT priming.  
 Directionally cloned. Size-selected on agarose gel,  
 average insert size 500 bp. Primary library,  
 non-amplified."

BASE COUNT 72 a 42 c 44 g 52 t

ORIGIN

Query Match 5.0%; Score 27; DB 45; Length 210;  
 Best Local Similarity 100.0%; Pred. No. 0.14;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 tcaatgaaaaaataaaaaa 542  
 |||||||||||||||||||||  
 Db 165 TCAATGAAAAAATAAAAAA 191

RESULT 15  
 AI675464/c  
 LOCUS AI675464 297 bp mRNA EST 19-MAY-1999  
 DEFINITION WB99g06.x1 NCI-CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:2313850 3',  
 mRNA sequence.  
 ACCESSION AI675464  
 VERSION AI675464.1 GI:4875944  
 KEYWORDS EST.  
 SOURCE

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 297)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On May 9, 1996 this sequence version replaced gi:1132969.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco.

Location/Qualifiers

1. .297  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:2313850"  
 /clone\_lib="NCI-CGAP\_Pr28"  
 /sex="male"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"

/note="organ: prostate; Vector: p773D-Pac (Pharmacia)  
 with a modified polylinker; Plasmid DNA from the  
 normalized library NCI-CGAP\_Pr22 was prepared, and ss  
 circles were made in vitro. Following HAP purification,  
 this DNA was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from a pool  
 of 5,000 clones made from the same library (cloneIDs  
 985608-986759, 110192-1101959, and 1217928-1220615).  
 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 94 a 46 c 73 g 84 t

ORIGIN

Query Match 5.0%; Score 27; DB 50; Length 297;

Best Local Similarity 100.0%; Pred. No. 0.13;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 tcaatgaaaaaataaaaaa 542

|||||||||||||||||||||

Db 29 TCAATGAAAAAATAAAAAA 3

Search completed: May 1, 2000, 14:44:27  
 Job time: 13987 sec





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:24:04 ; Search time 4425.31 Seconds  
(without alignments)  
-612.723 Million cell updates/sec

Title: US-09-215-435-113  
Perfect score: 893  
Sequence: 1 gccatgacgcacgtcac.....aaaaaaaaaaaaaaaaaaaaa 893

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 821193 seqs, -1518192014 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_bal.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vl.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vl.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_bal.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*

- 44: gb\_htg6.\*
- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	741	83.0	4035	40	AF160973 Homo sapi
2	741	83.0	4054	10	AB032994 Homo sapi
c 3	216	24.2	87184	41	AC009185 Homo sapi
4	50	5.6	1117	12	AF162472 Mus muscu
5	25	2.8	4112	12	AF072697 Mus muscu
6	24	2.7	689	3	AF030235 Oryctolag
7	24	2.7	4446	34	U72236 Dictyosteli
c 8	24	2.7	67157	44	AC016537 Homo sapi
c 9	24	2.7	109473	41	AC008317 Drosophill
10	24	2.7	169583	42	AC008578 Homo sapi
11	24	2.7	228224	42	AC008782 Homo sapi
12	23	2.6	1279	12	AF107299 Mus muscu
13	23	2.6	2316	12	AF107298 Mus muscu
14	23	2.6	6061	34	DWTF125 X98235 D.melanogas
c 15	23	2.6	172124	44	AC016509 Homo sapi
16	23	2.6	172161	42	AC012128 Homo sapi
c 17	22	2.5	461	13	G30354 human STS S
c 18	22	2.5	519	9	D43724 Human DNA f
19	22	2.5	1475	8	AF075582 Mesembrya
20	22	2.5	1863	11	HSU94357 Homo sapien
21	22	2.5	1938	10	HS8800122 AL049337 Homo sapi
22	22	2.5	2040	5	AR067642 Sequence
23	22	2.5	2110	7	V1702 V.faba mRNA
24	22	2.5	2685	10	HSU78294 U78294 Homo sapien
25	22	2.5	4379	9	HUMHA1025A D38549 Human mRNA
26	22	2.5	20014	7	SPCC31H12 AF031824 S.pombe c
27	22	2.5	28021	35	CEHX37E11B AF125451 Caenorhab
28	22	2.5	32000	8	SPCC1183 AL031740 S.pombe c
29	22	2.5	52597	11	HSJ421120 AL117327 Human DNA
c 30	22	2.5	72578	42	AC012432 Homo sapi
c 31	22	2.5	89779	7	AB005234 Arabidops
c 32	22	2.5	100000	10	AP000068 Homo sapi
c 33	22	2.5	102574	45	AC017935 Drosophill
34	22	2.5	103443	8	AC010164 Arabidops
c 35	22	2.5	114130	42	AC010659 Drosophill
c 36	22	2.5	114753	43	AC008356 Drosophill
c 37	22	2.5	120848	33	AC007851 Drosophill
38	22	2.5	132596	43	AC013046 Drosophill
c 39	22	2.5	140775	33	AC007173 Drosophill
c 40	22	2.5	143369	40	AC006316 Homo sapi
c 41	22	2.5	151720	33	AC008184 Drosophill
c 42	22	2.5	152409	32	PFMAL1P1 AL031744 Plasmodi
43	22	2.5	153400	42	AC009700 Homo sapi
c 44	22	2.5	155880	33	AC006739 Caenorhab
45	22	2.5	160386	44	AC016446 Homo sapi

ALIGNMENTS

RESULT 1  
AF160973  
LOCUS AF160973 4035 bp mRNA  
DEFINITION Homo sapiens p53 inducible protein (PIR121) mRNA, complete cds.  
ACCESSION AF160973  
VERSION AF160973.1 GI:5616319



```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hk08280"
/dev_stage="adult"
/sex="male"
/tissue_type="brain"
1..3847
/gene="KIAA1168"
<1..3847
/gene="KIAA1168"
/translation="APARGPAAVSGSPGPAAPVONTETAAMTHVTLDELALSNVDL
EGHYAVMLYTRWSCRAIPQVKNQEPNVEIYKIVLEPEVTKLMKPMYEMLE
IERFCVSEKCHAEKRDYSEAYLLTGRFINNFVLEKMKKSVKNDHSAKRR
AAQLRKMDPQSQESQNSFLANHNRIQCILHQLEVPVPGVEELADVNICVDY
YENKWLTPSPKMLKVMGFLYLDGNSVIYKLDKRNLSKIDKPKQLQVVP
LFGDMQELARYIETSAHYENKSKWTQSSISPOYNICQWQIIRDDHIFISELA
RYSEVVTGSLDSQSDSEYRELFLALRGLQLLKKSHHYMEVYSKVLHFTDFE
CNKCPGTAETREARYNTYSEKFAFVEVIAMIKGLQVLMGRNESVFNQAIINTIY
AALQDFALVRLQRAVRKKKNVLSVLAQIRKTCIDWEGREPPNDPCIRGEKP
KGGFDIKVPRVAVGPTSLQVMTLESLIADKSGSKTKLRSLLDGPVILAIEDFHK
QSFFTLNLNISEALQCCDLQSLWREFFLELTMGRRIQFIEMSPWILTDHLET
KEPSMVEVLPDLNDISAYALTCKFKFLYDEIAEAVNLCDQVFKIADQIFAV
YKAMAGSVLLDKRAECKNYGVIPIPPSNRYETLLKQHVQLLSRIDLNRLLTOR
ISAMYSLSQDAISRFSEDETSIVLEWLEINRLRHLRLLCKHMTLDSFDMFREAN
HNVSAPGRILTHVWELNFDLPNYCYNGSTNFRVTAIFTEQEPQDKPANKYQPY
LYGKPLNIAYSHIYSYRNVGPPHFKTCILGLYQGIAYVMEELKIVKSLQGTI
LQYVKTILNEMPKICRILPRHEYGSPGLEFFHQLKDIIEYAEKLTDFVQSLREVGNA
ILFCLLILQALSOEVCDDLHAAPFONILPRVIKEGERLEVRMKRLEAKYAPLHVP
LJERLGPQQAIAREGDLTKERLCCGLSMEVILPRISYLDQPIWRGPPPTNGVM
HYDECEPHRWSAMQFVYICIPVGNTEFTACQCFDGLNWWGCSILVLLGQRRFDLF
DFCYHLLKQVQDKGDELIKVNPLKMDRIRKYQIOLNNEVFALINKYMSVETDSS
VEHVRCPQPIHQSLATTC"
BASE COUNT 1003 a 1127 c 1052 g 872 t
ORIGIN

Query Match 83.0%; Score 741; DB 10; Length 4054;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 791; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 cagcagctcaccgtgaagatgcctgtccaaagtgagcctgacgtgaagagcttccct 68
DB 91 CACGCGCTCACCCGGAAGATGCCCTGTCCACAGTGGACCTGCTGAAGAGCTTCCCT 150
QY 69 ccccgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 128
DB 151 CCCCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 210
QY 129 tgacacaaactttgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 188
DB 211 TGACACAAACTTTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 270
QY 189 gctacagctccactcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 248
DB 271 GCTACAGTCCACTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 330
QY 249 gcttacacctgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 308
DB 331 GCTGTACACTGGCGAGCTGTTCCGGGCCATTTCCCGAGGTGAATGCAACAGCAGCC 390
QY 309 caaccagtagagatctatgagaagacagtagaggtgctggagccggaggtcaccagct 368
DB 391 CAACCGAGTAGAGATCTATGAGAGACAGTAGAGGTGCTGGAGCGCGGAGGTCAACG 450

```

```

QY 369 catgaagttcatgtattttcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 428
DB 451 CATGAAGTTCAATGATTTTTCAGCGCAAGGCCATCGAGCGGTCTTCGACGCGAGGTGAAGCG 510
QY 429 gctgtgcatcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 488
DB 511 GCTGTGCTCATGCCAGCGCAGGAGGACTTTGCTCTGAGCGCTACCTCTCTGAGCCTTGG 570
QY 489 caagtcatcaaatgtttgtctctctgtgagcagcagcagcagcagcagcagcagcagcagcagc 548
DB 571 CAAGTTCAATCAATGATTTTTCAGCGCAAGGCCATCGAGCGGTCTTCGACGCGAGGTGAAGCG 630
QY 549 gaatgacacactccgcctacaagagcagcagcagcagcagcagcagcagcagcagcagcagc 608
DB 631 GAATGACCACTCTGCTCTACAAGAGGCGACAGCTTCTCTGCGGAAGATGGCAGATCCCA 690
QY 609 gtcctacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 668
DB 691 GTCATCAGGAGTGCAGAGACCTTTCATGTTCTTCTGCGCAACACACAGGATCACCA 750
QY 669 gtgtctcaccagcaactgaagtgatccccaggctatgagcagcagcagcagcagcagcagcagc 728
DB 751 GTGTCTCACCAGCAACTTGAAGTGATCCAGGCTATGAGGAGCTGCTGGCTGACATTGT 810
QY 729 caacatctgtgattactacgagcaacacacacacacacacacacacacacacacacacacac 788
DB 811 CAACATCTGTGTGATTACTACGAGACAAGATGATGATGATGATGATGATGATGATGATGATGAT 870
QY 789 gctcctcaaggt 800
DB 871 GCTCCTCAAGGT 802

RESULT 3
AC009185/c
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CIT-HSPC_248019, *** SEQUENCING IN
PROGRESS ***, 2 ordered pieces.
AC009185
AC009185.2 GI:6015313
VERSION
KEYWORDS HTG; HTGS_PHASE2.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 87184)
AUTHORS
DOE Joint Genome Institute.
TITLE
Sequencing of Human Chromosome 5
JOURNAL
Unpublished
2 (bases 1 to 87184)
AUTHORS
DOE Joint Genome Institute.
TITLE
Direct Submission
JOURNAL
Submitted (05-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Oct 7, 1999 this sequence version replaced gi:5705980.
www.jgi.doe.gov.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved
* 1 14428: contig of 14428 bp in length
* 14429 gap of unknown length
* 14429 87184: contig of 72756 bp in length.
* Location/Qualifiers
* 1..87184
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="5"

```

```

Best Local Similarity 100.0%; Pred. No. 2.6e-15;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 gagcacaggaatgcatgttcacggcgcatgcaaggtacatgagcagc 191
|||||
Db 154 GAGGACAGCAATGCTTCTCAGCGGCATGCAAGGTACATTGAGCAGC 203

RESULT 5
AF072697 4112 bp mRNA ROD 08-JUL-1998
LOCUS Mus musculus SHYC (Shyc) mRNA, complete cds.
DEFINITION AF072697
ACCESSION AF072697.1 GI:3293550
VERSION AF072697.1
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4112)
AUTHORS Koester, F., Schinke, B., Niemann, S. and Hermans-Borgmeyer, I.
TITLE Identification of Shyc, a novel gene expressed in the murine
developing and adult nervous system
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 4112)
TITLE Koester, F., Schinke, B., Niemann, S. and Hermans-Borgmeyer, I.
JOURNAL Direct Submission
TITLE Submitted (16-JUN-1998) University of Hamburg, Center for Molecular
Neurobiology, Martinistr. 52, Hamburg D-20246, Germany
FEATURES
Location/Qualifiers
Source
1..4112
/organism="Mus musculus"
/strain="C57Bl/J6"
/db_xref="taxon:10090"
/tissue_type="brain"
gene 1..4112
CDS 34..3795
/genes="Shyc"
/notes="expressed in the nervous system during development
and in the adult brain"
/codon_start=1
/product="SHYC"
/protein_id="AAC25773.1"
/db_xref="GI:3293551"
/translation="MAAQVTLSDLSNVDLLEELPLDQPCIEPPSSLLYQPNFT
NFEDRNAFTGIATYEQATVHSSNMLEGEQYAVMLYTWRSRAIPQVKNEQ
NRVEIYKTVLEPEVTKLMNMYFORNATERGCEVRRLCHAEKRDVSEAYLLT
LGKTFNMFGLDELKMKCSVKNDHSAVKRAQFLKRMADPOSIOESQNLSELANHN
KITQSQOOLEVWISYEELLADIVNLCVDYENRMYLTPSEKMLLKVAGGLYMDG
SVSNYIKLDAKRINLSKIDKIFKQLOVPLFGDMQIELARYIKTSAHYEENKSWTC
ASSSSPOYNICEQMIQIREDHMFISELARYSNSEVTVGSGRQAKTDAEYRLFD
LALQGLQLSQSAHAMEVYSWKLVHPDYSKDCPDNAEYERATRYNTTEKFA
LVEVTAMIKGLVLMGRVESVNHAIHTVYAAQDFSQVTLREPLRAIKKKNVIO
SVLOAIRKTVCDWETGHEPNDPALRGKDPKSGFDIKVPRVAVGSPSTQLYMVTML
ESLIADKSGSKTIRSSLEGTLIDIEKFRSEFFYHLINSEFSELOQCDSQWFR
EFFLELTMRRIQPIEMSPWILLDIELETKREASMEYVLSLDLYNDSAHYATKF
NKQFLYDEIEAEVNLQFQFYKLDADIFAYIKVMAGSLDLRLRSEKQKQATFHL
PPSNRYETLLKQHVQLLGRSIDNRLITQVSAAMYKSELAIGRFESEDLTSVVEL
DGLLEINRHTKLLSRYLTLDSFDAMFEANHNVSAPYRITLHVFWELNDFLPNYC
YNGSTNRVRVLPSPQEFQKDPKQNAOPVHLGSKALNLAYSYIGSYRNFVGPFF
OVICELLYQGIADVMEELLKVKVSLQGTILQYVKTLMVMPKICRLPREYSGPI
LEFFHHQKDAEYAEKTVCFQNLREVGNAVFLCLLEEQSLSELEVCDLLHAAPFQ
ILPRTHQKGERVDAMKRLSKYAPLHLVFLIERLGTPOQIATAREGHLITKRLCC
GLTFVEITLIRTFLLDDPIWRGPLSPNGMVDECFEHLWSAMQVYCTPVGTHE
GTSVECFDGLHWAGCMIIVLGQORRAVLDFCVHLLKVKQHDGKDIKINVLKMK
VERIRKFOILNDEIITLIDKLYKSGDGTTPVEHVRFCQFPPIHQSSASS"
BASE COUNT 1126 a 986 c 1007 g 993 t
ORIGIN

Query Match 2.8%; Score 25; DB 12; Length 4112;
Best Local Similarity 100.0%; Pred. No. 0.03;

BASE COUNT 25707 a 18905 c 18505 g 24067 t
ORIGIN

Query Match 24.2%; Score 216; DB 41; Length 87184;
Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 667 cagtgtccaccagcaactgaagtatcccgaggtatgagagctgctggctgacatt 726
|||||
Db 9111 CAGTGTCTCCACCAGCAACTGAAGTGATCCAGGCTATGAGGAGCTGCTGGCTGACATT 9052

QY 727 gtaacaactctgtgtgattactacgagacaagaatgtacctgactccacagtgagaacat 786
|||||
Db 9051 GTCAACATCTGTGTGATTACTACGAGACAAGATGTACCTGACTCCACGTGAGAAACAT 8992

QY 787 atgtctctcaagtataaactccctcagcgccgaccatgagcctgagccttaccctctc 846
|||||
Db 8991 ATGCTCTCAAGGTAAACATCCCTGAGCGCCGACCACCTGAGCGCTTACCCCTCTC 8932

QY 847 acctctcttattaaaaatccgttttaaaaaacaa 882
|||||
Db 8931 ACCTCTCTCTATTAAAAATCCGTTTTAAAAACAA 8896

RESULT 4
AF162472 1117 bp mRNA ROD 28-JUL-1999
LOCUS Mus musculus Inducible protein (Pir121) mRNA, partial cds.
DEFINITION AF162472
ACCESSION AF162472
VERSION AF162472.1 GI:5616489
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1117)
AUTHORS Saller, E., Tom, E., Brunori, M., Otter, M., Estreicher, A., Mack, D. and
1990.R.
TITLE Increased apoptosis induction by 121f mutant p53
JOURNAL EMBO J. (1999) In press
REFERENCE 2 (bases 1 to 1117)
AUTHORS Saller, E. and Iggo, R.
TITLE Direct Submission
JOURNAL Submitted (24-JUN-1999) Oncogene Group, Swiss Institute for
Experimental Cancer Research, Ch. des Boveresses 155, Epalinges
CH-1066, Switzerland
FEATURES
Location/Qualifiers
Source
1..1117
/organism="Mus musculus"
/db_xref="taxon:10090"
gene 1..1117
CDS 16..>1117
/genes="Pir121"
/notes="Pir121"
/codon_start=1
/product="inducible protein"
/protein_id="AAD45803.1"
/db_xref="GI:5616490"
/translation="MTHTVTLSDLSNVDLLEELPLDQPCIEPPSSIMYQANFDT
NFEDRNAFTGIATYEQATVHSSNMLEEGHDYAVMLYTWRSRAIPQVKNEQ
NRVEIYKTVLEPEVTKLMNMYFORNATERGCEVRRLCHAEKRDVSEAYLLT
LGKTFNMFGLDELKMKCSVKNDHSAVKRAQFLKRMADPOSIOESQNLSELANHN
RITQCHQOOLEVPGYEELLADIVNICVDYENKMYLTPSEKMLLKVAGGLYMDG
NVSNIYKLDARINLSKIDKFFKQLOVPLFGDMQIELARYIETSAHYEENKSWTC
TOSSISPOYNICEQMIQIRDHIRFISELARYS"
BASE COUNT 296 a 295 c 295 g 231 t
ORIGIN

Query Match 5.6%; Score 50; DB 12; Length 1117;
```

















Institute, 44 Binney Street, Boston, MA 02115, USA					
FEATURES	Location/Qualifiers				
source	1..1279				
	/organism="Mus musculus"				
	/strain="ICK"				
	/db_xref="taxon:10090"				
gene	<1..1279				
	/gene="Tcf4"				
CDS	<1..1927				
	/note="alternatively spliced"				
	/codon_start=1				
	/product="HMG-box transcription factor TCF4E"				
	/protein_id="A016968.1"				
	/db_xref="GI:4324633"				
	/translation="QESSQSVDGSLHSSKHODSKKEEKKKPHIKKPLNFMFLMKRM RAKVVAECTLKESAINIILGRWHALSREQAKEYELARKEROLHMOLYPGWSARDN YGKKKKRRDQPGFTNDLSAPKCRARFFGLDQNNMCPADATPKCKRALFGLDRQ TLWCKPCRKKKCVRYIQEGSCSLSPSSDGLSDSPSPSHLLSGPPOAKSQOTQT QPLSLKPDPLAHLSMWPPPPALLAEAHGKASALCPNGALDLPPALQPQSNVFS SLAQSTSLSHNSLAGTPQLJUSLTKLE"				
BASE COUNT	342 a 357 c 285 g 295 t				
ORIGIN					
	Query Match 2.6%; Score 23; DB 12; Length 1279;				
	Best Local Similarity 100.0%; Pred. No. 0.38; Mismatches 0; Indels 0; Gaps 0;				
MATCHES	23; Conservative 0; Indels 0; Gaps 0;				
QY	871 tttaaaaaacaaaaaaaaaa 893				
DB	1255 TTTAAAAACAAAAAAAAAA 1277				
RESULT 13					
AF107298	2316 bp mRNA	ROD	03-MAR-1999		
LOCUS	Mus musculus HMG-box transcription factor TCF4B (Tcf4) mRNA,				
DEFINITION	complete cds.				
VERSION	AF107298				
KEYWORDS	AF107298.1 GI:4324630				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Lee,Y.J., Swencki,B., Shoitchet,S. and Shivdasani,R.A.				
TITLE	A possible role for the high mobility group box transcription factor Tcf-4 in vertebrate gut epithelial cell differentiation				
JOURNAL	J. Biol. Chem. 274 (3), 1566-1572 (1999)				
MEDLINE	99098900				
REFERENCE	2 (bases 1 to 2316)				
AUTHORS	Lee,Y.J. and Shivdasani,R.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-NOV-1998) Adult Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA				
FEATURES	Location/Qualifiers				
source	1..2316				
	/organism="Mus musculus"				
	/strain="ICK"				
	/db_xref="taxon:10090"				
gene	1..2316				
	/gene="Tcf4"				
CDS	277..1620				
	/note="alternatively spliced"				
	/codon_start=1				
	/product="HMG-box transcription factor TCF4B"				
	/protein_id="A016967.1"				
	/db_xref="GI:4324631"				
	/translation="MPQLNGGGDDLGANDLISFKDEQFEKKNSENSAERLDADV KSSLVNSETNQNSSEAREPPRPSFRDKSRLESLEAAKRQDGGLFGPPYFG				

LVMPGDATLQIONNPATSLMTPOGVMLETRTPSPONKPFETSPSAGGQOYLVGANGOL  
 SPIQIISTPMLGWPQOQGGAFVQASPTITITIQQPPAPQTIQISQPAQATM  
 TETVMSQGTAAAPDPTTCTSPRSPSSHRSGSDWQCVSSSEPDAAVSPQSTES  
 RPSFSDICSGIKNVFTQPSGIIYKHSEPKIRIHIITQTSIAENISLMQOGOMRL  
 ATSNSTGASMSITSSSAPLTTTALAGQRLMHRQQAHHHPNISNSHPNPTFLHS  
 HKVNFYQGGDPSLTSTNCITHTLTHQHTTFLATA"

polyA\_signal  
 6026...6031  
 polyA\_site  
 6036  
 BASE COUNT 1721 a 1727 c 1429 g 1184 t  
 ORIGIN

Query Match 2.6%; Score 23; DB 34; Length 6061;

Best Local Similarity 100.0%; Pred. No. 0.32; Mismatches 0; Indels 0; Gaps 0;

Qy 871 ttttaaaacacaaaaaa 893  
 |||||  
 Db 6034 TTTAAAAACAAAAAAA 6056

# RESULT 15

AC016509/c  
 LOCUS AC016509 72124 bp DNA HTG 01-DEC-1999  
 DEFINITION Homo sapiens clone RP11-10G6, LOW-PASS SEQUENCE SAMPLING.  
 ACCESSION AC016509  
 VERSION AC016509.1 GI:6492517  
 KEYWORDS HTG; HTGS-PHASE0.  
 SOURCE human.

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens, clone RP11-10G6

## REFERENCE

1 (bases 1 to 72124)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Baldwin,J., Barna,N., Beckerly,R., Boguski,K., Boulikas,G.,  
 Brown,A., Castle,A., Collins,S., Collins,S., Collins,S., Collins,S.,  
 Cooke,P., DeRubeis,K., Dewar,K., Domingue,M., Donnelly,L., Doyle,M.,  
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
 Galagan,J., Gargyala,S., Grant,G., Hagan,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Lehotzky,J., Liu,C., Locke,K., MacDonald,P., Marquis,N.,  
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,  
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.  
 Direct Submision  
 Submitted (01-DEC-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

## TITLE

JOURNAL  
 COMMENT  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submission@genome.wi.mit.edu  
 Project Information  
 Center project name: L3066  
 Center clone name: 10\_G\_6

\* NOTE: This record contains 82 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.

\* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 859: contig of 859 bp in length  
 \* gap of unknown length  
 \* 860 1751: contig of 892 bp in length  
 \* gap of unknown length  
 \* 1752 2628: contig of 877 bp in length  
 \* gap of unknown length  
 \* 2629 3496: contig of 868 bp in length  
 \* gap of unknown length  
 \* 3497 4387: contig of 891 bp in length  
 \* gap of unknown length  
 \* 4388 5245: contig of 858 bp in length  
 \* gap of unknown length  
 \* 5246 6134: contig of 889 bp in length  
 \* gap of unknown length  
 \* 6135 7000: contig of 866 bp in length  
 \* gap of unknown length  
 \* 7001 7884: contig of 884 bp in length  
 \* gap of unknown length  
 \* 7885 8753: contig of 869 bp in length  
 \* gap of unknown length  
 \* 8754 9629: contig of 876 bp in length  
 \* gap of unknown length  
 \* 9630 10505: contig of 876 bp in length  
 \* gap of unknown length  
 \* 10506 11386: contig of 881 bp in length  
 \* gap of unknown length  
 \* 11387 12265: contig of 879 bp in length  
 \* gap of unknown length  
 \* 12266 13155: contig of 890 bp in length  
 \* gap of unknown length  
 \* 13156 14017: contig of 862 bp in length  
 \* gap of unknown length  
 \* 14018 14885: contig of 868 bp in length  
 \* gap of unknown length  
 \* 14886 15776: contig of 891 bp in length  
 \* gap of unknown length  
 \* 15777 16675: contig of 899 bp in length  
 \* gap of unknown length  
 \* 16676 17543: contig of 868 bp in length  
 \* gap of unknown length  
 \* 17544 18428: contig of 885 bp in length  
 \* gap of unknown length  
 \* 18429 19305: contig of 877 bp in length  
 \* gap of unknown length  
 \* 19306 20191: contig of 886 bp in length  
 \* gap of unknown length  
 \* 20192 21062: contig of 871 bp in length  
 \* gap of unknown length  
 \* 21063 21898: contig of 836 bp in length  
 \* gap of unknown length  
 \* 21899 22774: contig of 876 bp in length  
 \* gap of unknown length  
 \* 22775 23643: contig of 869 bp in length  
 \* gap of unknown length  
 \* 23644 24570: contig of 927 bp in length  
 \* gap of unknown length  
 \* 24571 25439: contig of 869 bp in length  
 \* gap of unknown length  
 \* 25440 26316: contig of 877 bp in length  
 \* gap of unknown length  
 \* 26317 27191: contig of 875 bp in length  
 \* gap of unknown length  
 \* 27192 28070: contig of 879 bp in length  
 \* gap of unknown length  
 \* 28071 28950: contig of 880 bp in length  
 \* gap of unknown length  
 \* 28951 29808: contig of 858 bp in length  
 \* gap of unknown length  
 \* 29809 30679: contig of 871 bp in length







**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 18:51:09 ; Search time 250.34 Seconds  
(without alignments)  
892.472 Million cell updates/sec

Title: US-09-215-435-113

Perfect score: 893

Sequence: 1 gcatgatcagcagctcac.....aaaaaaaaaaaaaaaaaaaaa 893

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 311585 seqs, 125096042 residues

Word size : 0

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	177	19.8	287	1 X51880	Human secreted pro
2	22	2.5	1863	1 V34332	Human glycerol-2
3	22	2.5	2040	1 V63088	Human tcf-4B cDNA.
4	21	2.4	539	1 Q03796	Trichinella spir
5	20	2.2	318	1 V21233	Homo sapiens clone
6	20	2.2	372	1 T92852	Flea saliva protei
7	20	2.2	372	1 V73400	Flea saliva protei
8	20	2.2	767	1 N82022	Gene encoding 25 k
9	20	2.2	767	1 Q43503	Sequence encoding
10	20	2.2	787	1 X37481	Human secreted pro
11	20	2.2	822	1 X51709	DNA encoding a hum
12	20	2.2	859	1 T62944	Human eosinocyte C
13	20	2.2	926	1 X22248	Human secreted pro
14	20	2.2	1126	1 X37441	Human secreted pro
15	20	2.2	1244	1 T67289	Human secreted pro
16	20	2.2	1448	1 X37462	Human secreted pro
17	20	2.2	1667	1 Q48589	Alcohol oxidase ge
18	20	2.2	1683	1 X37423	Human secreted pro
19	20	2.2	1747	1 T04201	Tobacco citrate sy
20	20	2.2	1798	1 T08611	Yeast heterogene e
21	20	2.2	1921	1 T08610	Yeast heterogene e
22	20	2.2	2065	1 V23062	BRCA1 modulator pr
23	20	2.2	2071	1 X27940	Human l(3)mbt prot
24	20	2.2	2250	1 V99806	Murine lysophospha
25	20	2.2	2400	1 Q42496	Full-length human
26	20	2.2	2445	1 N60111	Sequence encoding
27	20	2.2	2445	1 Q48393	Full-length foetal
28	20	2.2	2744	1 Q31769	Gastrin-binding pr
29	20	2.2	3311	1 T70377	Cytohesin 1. Cyth
30	20	2.2	3394	1 X27920	Human l(3)mbt prot
31	20	2.2	3410	1 T03887	Human mGLUR3 DNA.
32	20	2.2	3512	1 X27921	Human l(3)mbt prot
33	20	2.2	7397	1 V60578	Human tumour suppr

34 20 2.2 8601 1 Q24828 AFP-1. DNA encodin  
35 20 2.2 8601 1 Q24829 AFP-1 (C 7508 T).  
36 19 2.1 26 1 V12482 Oligonucleotide SE  
37 19 2.1 26 1 V59215 Circular template  
38 19 2.1 26 1 X30018 Precircle DNA olig  
39 19 2.1 29 1 V59216 Linear multimer pr  
40 19 2.1 39 1 V12483 Oligonucleotide SE  
41 19 2.1 39 1 X30019 Multimer SEQ ID NO  
42 19 2.1 106 1 V11611 Homo sapiens adult  
43 19 2.1 146 1 T20709 Human gene signatu  
44 19 2.1 955 1 X20415 Human secreted pro  
45 19 2.1 1089 1 Q94250 Cysteine proteinas

#### ALIGNMENTS

RESULT 1  
X51880  
ID X51880 standard; DNA; 287 BP.  
AC X51880;  
DT 22-JUN-1999 (first entry)  
DE Human secreted protein 5; EST SEQ ID NO: 94.  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulation; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
OS Homo sapiens.  
PN WO9906552-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; IB1236.  
PR 01-AUG-1997; US-905223.  
PA (GEST ) GENSET.  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
DR WPI: 99-153782/13.  
DR P-PSDB: Y13080.  
PT New isolated brain-derived nucleic acids - used to develop products  
PT which may have cytokine, immune, regulatory, haematopoiesis  
PT regulating, anti-inflammatory or tumour inhibition activity  
PS Claim 1; Page 244; 577pp; English.  
CC X51787 to X52019 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins, and encode the proteins given in Y12987 to Y13219,  
CC respectively. The proteins given represent the signal peptide and an  
CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
CC can be used for producing secreted human gene products. They can also  
CC be used to develop products for diagnosis and therapy. The proteins  
CC obtained may have cytokine activity, cell proliferation/differentiation  
CC activity, haematopoiesis regulating activity, tissue growth regulating  
CC activity, reproductive hormone regulating activity, chemotactic/  
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
CC or other activities. The products can be used in forensic, gene therapy  
CC and chromosome mapping procedures. The sequences can also be used for  
CC obtaining corresponding promoter sequences. The nucleic acids encoding  
CC the signal peptide can be used for directing extracellular secretion of  
CC a polypeptide or the insertion of a polypeptide into a membrane, or  
CC importing a polypeptide into a cell.  
SQ Sequence 287 BP; 66 A; 101 C; 71 G; 49 T;

Query Match 19.8%; Score 177; DB 1; Length 287;  
Best Local Similarity 100.0%; Pred. No. 3.1e-71;  
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 cagcgacgtcacccctggagaatgcctgtcccaactgagcctgttggaagagcttccct 68  
DB 103 CAGCGACGTCACCTGGGAAGATGCCCTGTCCAACGTGGACCTGCTTTGAAGAGCTTCCCT 162  
QY 69 ccccgaccagcagcatgcatgagcctccaccttctccatcatgtaccaggctaacct 128  
DB 163 CCCCAGACAGAGCCATGCGAGCTCCAGCTTCTCCATCATGTGTCAGGCTAACTT 222



Db 23 TAAAAACAAAAA 3

## RESULT 5

V21233 ID V21233 standard; cDNA; 318 BP.  
AC V21233;  
DT 03-AUG-1998 (first entry)  
DE Homo sapiens clone CC365\_40 3' end.  
KW secreted protein; human; adult brain; ds.  
OS Homo sapiens.  
PN W09807859-A2.  
PD 26-FEB-1998.  
PF 22-AUG-1997; U14874.  
PR 23-AUG-1996; US-702344.  
PA (GEM) Genetics Inst Inc.  
PI Merberg F, McCoy JM, Lavalie ER, Racie LA, Treacy M, Spaulding V, Jacobs K.  
PI WPI; 98-169163/15.  
DR New nucleic acid encoding secreted proteins from human cells - useful e.g. as immuno-modulators, antitumour agents, promoters of tissue growth, haemostatic and thrombolytic agents etc.  
PS Claim 24; Page 56; 79pp; English.  
CC The sequence is that of the 3' end of clone CC365\_40 (including the polyA tail) which was isolated from a human adult brain cDNA library using methods selective for cDNAs that encode secreted proteins.  
CC  
SQ Sequence 318 BP; 93 A; 38 C; 75 G; 100 T;

Query Match 2.2%; Score 20; DB 1; Length 318;

Best Local Similarity 100.0%; Pred. No. 4.5;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 874 aaaaaacaaaaa 893

DB 298 AAAAAACAAAAA 317

## RESULT 6

T92852 ID T92852 standard; cDNA; 372 BP.  
AC T92852;  
DT 14-APR-1998 (first entry)  
DE Flea saliva protein fspn(M) cDNA clone nfspn(M)372.  
KW Flea saliva protein; fspn(M); allergic dermatitis; therapy;  
KW diagnosis; nfspn(M)372; ss.  
OS Ctenocephalides felis.  
PN W09737676-A1.  
PD 16-OCT-1997.  
PF 10-APR-1997; U05959.  
PR 10-APR-1996; US-630822.  
PA (HESK-) HESKA CORP.  
PI Hunter SW, Sim G, Weber ER;  
PI WPI; 97-512409/47.  
DR New flea saliva proteins - useful for treating allergic dermatitis and as diagnostic reagents  
PS Example 5; Page 137-38; 179pp; English.  
CC This cDNA clone, designated nfspn(M)372, encodes a portion of flea saliva protein fspn(M). It was isolated from a flea salivary gland ZAP-cDNA library by immunoscreening with IgE enriched dog antiserum. Novel nucleic acids (see T92820-27) and novel flea saliva proteins (see W30480-91), their fragments or mimetopes can be used in claimed methods for treating allergic dermatitis in an animal, for testing if an animal is susceptible to, or has, allergic dermatitis and for desensitising a host animal to allergic dermatitis.  
CC  
SQ Sequence 372 BP; 133 A; 50 C; 61 G; 127 T;

Query Match

Best Local Similarity 2.2%; Score 20; DB 1; Length 372;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 874 aaaaaacaaaaa 893

DB 349 AAAAAACAAAAA 368

## RESULT 7

V73400 ID V73400 standard; cDNA; 372 BP.  
AC V73400;  
DT 26-APR-1999 (first entry)  
DE Flea saliva protein nucleic acid nfspn(M)372.  
KW Flea saliva protein; nfspn(M)372; allergic dermatitis; allergen;  
KW allergy; therapy; diagnosis; vaccine; ectoparasite; ss.  
OS Ctenocephalides sp.  
PN W09845408-A2.  
PD 15-OCT-1998.  
PF 15-OCT-1997; U18669.  
PR 10-APR-1997; W0-U05959.  
PA (HESK-) HESKA CORP.  
PI Weber ER;  
PI WPI; 98-594480/50.  
DR New isolated ectoparasite saliva genes - used to develop products for the diagnosis, prevention, treatment and determining susceptibility to allergic dermatitis  
PT  
PS Example 5; Page 107; 172pp; English.  
CC This is the nucleotide sequence of nfspn(M)372 comprising a portion of a flea saliva protein gene. nfspn(M)372 was isolated from a flea salivary gland cDNA expression library by immunoscreening with IgE-enriched canine antiserum. The invention is directed to novel products and methods for isolating ectoparasite saliva proteins (ESPs), including flea saliva proteins. It provides ESps (claimed, see W82382-93), nucleic acid molecules encoding them, methods for their recombinant production, CC therapeutic compositions for treating allergic dermatitis that comprise at least one ESP, as well as assay kits for testing if an animal has, or is susceptible to, allergic dermatitis, and a method of desensitising a host animal to allergic dermatitis using ESps. CC The ESps can also be used for the production of antibodies useful in diagnosis or in vaccines for passive immunisation against allergic dermatitis.  
CC  
SQ Sequence 372 BP; 133 A; 50 C; 61 G; 127 T;

Query Match 2.2%; Score 20; DB 1; Length 372;

Best Local Similarity 100.0%; Pred. No. 4.5;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 874 aaaaaacaaaaa 893

DB 349 AAAAAACAAAAA 368

## RESULT 8

N82022 ID N82022 standard; DNA; 767 BP.  
AC N82022;  
DT 22-OCT-1990 (first entry)  
DE Gene encoding 25 kD surface antigen, Pfs25.  
KW 25 kD surface antigen; Pfs25; vaccine; malaria; zygotes; ookinits;  
KW mosquito; transmission-blocking immunity; sexual development; ss.  
OS Plasmodium falciparum.  
FH Key Location/Qualifiers  
FT signal\_peptide 81..122 /tag= a  
FT mat\_peptide 123..725 /tag= b  
FT /label=pfs25

US7188918-A.

08-NOV-1988.

02-MAY-1988; 188918.

02-MAY-1988; US-188918.









```

FT      /*tag= g
FT      /note= "G is one clone, T in 2 others"
FT      1174
FT      /*tag= h
FT      /note= "one clone has polyA inserted"
PN      WO9722702-A1.
PD      26-JUN-1997.
PF      25-NOV-1996; U18859.
PR      18-DEC-1995; US-008809.
PR      14-AUG-1996; US-023943.
PA      (SUGEN-) SUGEN INC.
PI      Mossie KG, Plowman GD;
DR      WPI: 97-341693/31.
DR      P-PSDB; W18083.
PT      Aurora-1 and Aurora-2 and related genes - useful in tumour gene
PT      therapy
PS      Claim 1; Page 79; 98pp; English.
CC      2 cDNA clones (T67289 and T67290) respectively code for novel
CC      proteins, termed Aurora-1 (W18083) and Aurora-2 (W18084) (AUR-1
CC      and AUR-2), which are related serine/threonine kinases with short
CC      N-terminal extensions and which appear to be involved in cancer
CC      and/or signal transduction disorders. To isolate the clones,
CC      primers (see also T67291-92) based on the conserved residues
CC      within the kinase domain of CCK4 were used to identify novel
CC      kinases in HEPH (palatal mesenchyme) RNA, and an isolated
CC      sequence was used as a probe to screen human pancreatic cancer
CC      cDNA. Polymorphic sites were identified in pancreatic tumour and
CC      normal duodenum sequences of both AUR-1 and AUR-2 cDNAs. AUR
CC      nucleic acids can be used to produce Aurora polypeptides,
CC      as probes and in gene therapy of AUR-related disorders.
SQ      Sequence 1244 BP; 293 A; 351 C; 330 G; 270 T;

Query Match      2.2%; Score 20; DB 1; Length 1244;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      874 aaaaaaaaaaaaaaaaaa 893
        |||
Db      1223 AAAAAAAAAAAAAAAAAA 1242

```

Search completed: May 1, 2000, 18:51:15  
Job time: 18759 sec

**THIS PAGE BLANK (COPY 10)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:04:44 ; Search time 181.17 Seconds  
(without alignments)  
590.124 Million cell updates/sec

Title: US-09-215-435-113  
Perfect score: 893  
Sequence: 1 gcatgatcacgcacgtcac.....aaaaaaaaaaaaaaaaaaaaa 893

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 214294 seqs, 59861574 residues

Word size : 0

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PCTUS9\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	2.5	2040	3	US-08-821-355A-1
2	22	2.5	2040	4	US-09-003-687A-1
3	21	2.4	539	7	5422363-3
4	20	2.2	56	1	US-08-273-169-1
5	20	2.2	318	2	US-08-702-344-17
6	20	2.2	372	3	US-08-630-822A-96
7	20	2.2	372	4	US-09-005-069-96
8	20	2.2	737	4	US-08-691-814B-116
9	20	2.2	767	1	US-07-697-275-1
10	20	2.2	1244	4	US-08-755-728-1
11	20	2.2	1244	4	US-08-974-655-1
12	20	2.2	1667	2	US-08-485-284A-1
13	20	2.2	1798	3	US-08-557-128-12
14	20	2.2	1921	3	US-08-557-128-11
15	20	2.2	2065	4	US-08-968-751-1
16	20	2.2	2400	7	5215909-13
17	20	2.2	2445	7	5215909-9
18	19	2.1	1104	6	PCT-US93-05640-9
19	19	2.1	1227	2	US-08-080-386-1
20	19	2.1	1227	4	US-08-390-000A-1
21	19	2.1	2585	4	US-08-579-777A-1
22	19	2.1	4088	3	US-08-317-310A-1
23	19	2.1	4088	6	PCT-US95-13041-1
24	18	2.0	23	2	US-07-869-933-7
25	18	2.0	25	1	US-08-113-646A-42
26	18	2.0	197	1	US-08-392-678-5

Sequence 5, Appli  
Sequence 5, Appli  
Sequence 4, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 19, Appli  
Sequence 3, Appli  
Sequence 10, Appli  
Sequence 8, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 5, Appli  
Sequence 5, Appli

27 18 2.0 197 1 US-08-457-304A-5  
28 18 2.0 197 1 US-08-456-701A-5  
29 18 2.0 1127 3 US-08-392-546C-4  
30 18 2.0 1476 1 US-08-075-533-3  
31 18 2.0 1476 4 US-08-948-176-3  
32 18 2.0 1476 6 PCT-US91-09160-3  
33 18 2.0 1519 1 US-07-971-759-19  
34 18 2.0 1680 4 US-08-700-152A-3  
35 18 2.0 2294 4 US-08-967-508-10  
36 18 2.0 2294 6 PCT-US94-02552-10  
37 18 2.0 2791 4 US-09-016-000-8  
38 18 2.0 2834 1 US-08-471-251A-1  
39 18 2.0 3207 1 US-07-946-497-1  
40 18 2.0 3207 2 US-08-483-322-1  
41 18 2.0 3207 3 US-08-478-882-1  
42 18 2.0 5630 4 US-08-937-931-1  
43 18 2.0 6008 1 US-07-789-915A-5  
44 18 2.0 6008 1 US-08-005-002C-5  
45 18 2.0 6008 1 US-08-487-203A-5

#### ALIGNMENTS

RESULT 1  
US-08-821-355A-1  
; Sequence 1, Application US/08821355A  
; Patent No. 5851775  
; GENERAL INFORMATION:  
; APPLICANT: Barker, Nick  
; APPLICANT: Clevers, Hans  
; APPLICANT: Korinek, Vladimir  
; APPLICANT: Morin, Patrice  
; APPLICANT: Kinzler, Kenneth  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Sparks, Andrew  
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC  
; TITLE OF INVENTION: Interact to Prevent Cancer  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: 1001 G Street, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/821.355A  
; FILING DATE: 20-MAR-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A  
; REGISTRATION NUMBER: 32,145  
; REFERENCE/DOCKET NUMBER: 1107.05064  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; TELEX: 97430 BMB UT  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2040 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-821-355A-1



```

> ZIP: 80203
> COMPUTER READABLE FORM:
> MEDIUM TYPE: Floppy disk
> COMPUTER: IBM PC compatible
> OPERATING SYSTEM: PC-DOS/MS-DOS
> SOFTWARE: Patentin Release #1.0, Version #1.25
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/08/630,822A
> FILING DATE: 11-APR-1996
> CLASSIFICATION: 435
> ATTORNEY/AGENT INFORMATION:
> NAME: CONNELL, GARY J.
> REGISTRATION NUMBER: 32,020
> REFERENCE/DOCKET NUMBER: 2618-17-C3
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: (303) 863-9700
> TELEFAX: (303) 863-0223
> INFORMATION FOR SEQ ID NO: 96:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 372 base pairs
> TYPE: nucleic acid
> STRANDEDNESS: single
> TOPOLOGY: linear
> MOLECULE TYPE: cDNA
> US-08-630-822A-96

Query Match 2.2%; Score 20; DB 3; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.5; Indels
Matches 20; Conservative 0; Mismatches 0;

Qy 874 aaaaacaaaaaa 893
      |||iiiiiiiii
Db 349 AAAAAACAAAAAA 368

RESULT 7
US-09-005-069-96
> Sequence 96, Application US/09005069
> Patent No. 5932470
> GENERAL INFORMATION:
> APPLICANT: FRANK, GLENN R.
> APPLICANT: HUNTER, SHIRLEY WU
> APPLICANT: WALLENFELS, LYNDA
> TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
> TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
> NUMBER OF SEQUENCES: 107
> CORRESPONDENCE ADDRESS:
> ADDRESSEE: Sheridan Ross P.C.
> STREET: 1700 Lincoln Street, Suite 3500
> CITY: Denver
> STATE: Colorado
> COUNTRY: U.S.A.
> ZIP: 80203
> COMPUTER READABLE FORM:
> MEDIUM TYPE: Floppy disk
> COMPUTER: IBM PC compatible
> OPERATING SYSTEM: PC-DOS/MS-DOS
> SOFTWARE: Patentin Release #1.0, Version #1.25
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/09/005,069
> FILING DATE:
> CLASSIFICATION:
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: 08/630,822
> FILING DATE: 11-APR-1996
> ATTORNEY/AGENT INFORMATION:
> NAME: CONNELL, GARY J.
> REGISTRATION NUMBER: 32,020
> REFERENCE/DOCKET NUMBER: 2618-17-C3
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: (303) 863-9700
> TELEFAX: (303) 863-0223

```





RESULT 14  
US-08-557-128-11  
; Sequence 11, Application US/08557128  
; Patent No. 5849524  
; GENERAL INFORMATION:  
; APPLICANT: KONDO, Keiji  
; APPLICANT: KAJIWARA, Susumu  
; APPLICANT: MISAHA, No. 5849524ihiko  
; TITLE OF INVENTION: TRANSFORMATION SYSTEMS FOR THE YEAST  
; TITLE OF INVENTION: CANDIDA UTILIS AND THE EXPRESSION OF  
; TITLE OF INVENTION: THEREWITH  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/557,128  
; FILING DATE: 25-JAN-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/JP95/01005  
; FILING DATE: 25-MAY-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 7-129287  
; FILING DATE: 28-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-285823  
; FILING DATE: 26-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-135015



NAME/KEY: CDS  
LOCATION: 103..1512



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 19:59:41 ; Search time 4088.29 Seconds  
(without alignments)  
-749.040 Million cell updates/sec

Title: US-09-215-435-113  
Perfect score: 893  
Sequence: 1 gcatatgcacgcgtcac.....aaaaaaaaaaaaaaaaaaaaa 893

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 5265378 seqs, -1714608768 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 10530736

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Pending Patents,NA:\*  
1: /cgn2\_6/ptodata/1/pna/PCTUS9\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US081A\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US081B\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US081C\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US081D\_COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US082A\_COMB.seq.\*  
9: /cgn2\_6/ptodata/1/pna/US082B\_COMB.seq.\*  
10: /cgn2\_6/ptodata/1/pna/US082C\_COMB.seq.\*  
11: /cgn2\_6/ptodata/1/pna/US083A\_COMB.seq.\*  
12: /cgn2\_6/ptodata/1/pna/US083B\_COMB.seq.\*  
13: /cgn2\_6/ptodata/1/pna/US084A\_COMB.seq.\*  
14: /cgn2\_6/ptodata/1/pna/US084B\_COMB.seq.\*  
15: /cgn2\_6/ptodata/1/pna/US084C\_COMB.seq.\*  
16: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*  
17: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*  
18: /cgn2\_6/ptodata/1/pna/US087A\_COMB.seq.\*  
19: /cgn2\_6/ptodata/1/pna/US087B\_COMB.seq.\*  
20: /cgn2\_6/ptodata/1/pna/US087C\_COMB.seq.\*  
21: /cgn2\_6/ptodata/1/pna/US088A\_COMB.seq.\*  
22: /cgn2\_6/ptodata/1/pna/US088B\_COMB.seq.\*  
23: /cgn2\_6/ptodata/1/pna/US088C\_COMB.seq.\*  
24: /cgn2\_6/ptodata/1/pna/US088D\_COMB.seq.\*  
25: /cgn2\_6/ptodata/1/pna/US089A\_COMB.seq.\*  
26: /cgn2\_6/ptodata/1/pna/US089B\_COMB.seq.\*  
27: /cgn2\_6/ptodata/1/pna/US089C\_COMB.seq.\*  
28: /cgn2\_6/ptodata/1/pna/US089D\_COMB.seq.\*  
29: /cgn2\_6/ptodata/1/pna/US089E\_COMB.seq.\*  
30: /cgn2\_6/ptodata/1/pna/US089F\_COMB.seq.\*  
31: /cgn2\_6/ptodata/1/pna/US089G\_COMB.seq.\*  
32: /cgn2\_6/ptodata/1/pna/US089H\_COMB.seq.\*  
33: /cgn2\_6/ptodata/1/pna/US089I\_COMB.seq.\*  
34: /cgn2\_6/ptodata/1/pna/US090A\_COMB.seq.\*  
35: /cgn2\_6/ptodata/1/pna/US090B\_COMB.seq.\*  
36: /cgn2\_6/ptodata/1/pna/US090C\_COMB.seq.\*  
37: /cgn2\_6/ptodata/1/pna/US090D\_COMB.seq.\*  
38: /cgn2\_6/ptodata/1/pna/US090E\_COMB.seq.\*  
39: /cgn2\_6/ptodata/1/pna/US091A\_COMB.seq.\*  
40: /cgn2\_6/ptodata/1/pna/US091B\_COMB.seq.\*  
41: /cgn2\_6/ptodata/1/pna/US091C\_COMB.seq.\*  
42: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq.\*  
43: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq.\*

44: /cgn2\_6/ptodata/1/pna/US092C\_COMB.seq.\*  
45: /cgn2\_6/ptodata/1/pna/US092D\_COMB.seq.\*  
46: /cgn2\_6/ptodata/1/pna/US092E\_COMB.seq.\*  
47: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq.\*  
48: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq.\*  
49: /cgn2\_6/ptodata/1/pna/US093C\_COMB.seq.\*  
50: /cgn2\_6/ptodata/1/pna/US093D\_COMB.seq.\*  
51: /cgn2\_6/ptodata/1/pna/US093E\_COMB.seq.\*  
52: /cgn2\_6/ptodata/1/pna/US093F\_COMB.seq.\*  
53: /cgn2\_6/ptodata/1/pna/US094A\_COMB.seq.\*  
54: /cgn2\_6/ptodata/1/pna/US094B\_COMB.seq.\*  
55: /cgn2\_6/ptodata/1/pna/US094C\_COMB.seq.\*  
56: /cgn2\_6/ptodata/1/pna/US094D\_COMB.seq.\*  
57: /cgn2\_6/ptodata/1/pna/US094E\_COMB.seq.\*  
58: /cgn2\_6/ptodata/1/pna/US094F\_COMB.seq.\*  
59: /cgn2\_6/ptodata/1/pna/US094G\_COMB.seq.\*  
60: /cgn2\_6/ptodata/1/pna/US094H\_COMB.seq.\*  
61: /cgn2\_6/ptodata/1/pna/US094I\_COMB.seq.\*  
62: /cgn2\_6/ptodata/1/pna/US094J\_COMB.seq.\*  
63: /cgn2\_6/ptodata/1/pna/US094K\_COMB.seq.\*  
64: /cgn2\_6/ptodata/1/pna/US094L\_COMB.seq.\*  
65: /cgn2\_6/ptodata/1/pna/US094M\_COMB.seq.\*  
66: /cgn2\_6/ptodata/1/pna/US094N\_COMB.seq.\*  
67: /cgn2\_6/ptodata/1/pna/US094O\_COMB.seq.\*  
68: /cgn2\_6/ptodata/1/pna/US094P\_COMB.seq.\*  
69: /cgn2\_6/ptodata/1/pna/US094Q\_COMB.seq.\*  
70: /cgn2\_6/ptodata/1/pna/US094R\_COMB.seq.\*  
71: /cgn2\_6/ptodata/1/pna/US094S\_COMB.seq.\*  
72: /cgn2\_6/ptodata/1/pna/US094T\_COMB.seq.\*  
73: /cgn2\_6/ptodata/1/pna/US094U\_COMB.seq.\*  
74: /cgn2\_6/ptodata/1/pna/US094V\_COMB.seq.\*  
75: /cgn2\_6/ptodata/1/pna/US094W\_COMB.seq.\*  
76: /cgn2\_6/ptodata/1/pna/US094X\_COMB.seq.\*  
77: /cgn2\_6/ptodata/1/pna/US094Y\_COMB.seq.\*  
78: /cgn2\_6/ptodata/1/pna/US094Z\_COMB.seq.\*  
79: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq.\*  
80: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq.\*  
81: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq.\*  
82: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq.\*  
83: /cgn2\_6/ptodata/1/pna/US095E\_COMB.seq.\*  
84: /cgn2\_6/ptodata/1/pna/US095F\_COMB.seq.\*  
85: /cgn2\_6/ptodata/1/pna/US095G\_COMB.seq.\*  
86: /cgn2\_6/ptodata/1/pna/US095H\_COMB.seq.\*  
87: /cgn2\_6/ptodata/1/pna/US095I\_COMB.seq.\*  
88: /cgn2\_6/ptodata/1/pna/US095J\_COMB.seq.\*  
89: /cgn2\_6/ptodata/1/pna/US095K\_COMB.seq.\*  
90: /cgn2\_6/ptodata/1/pna/US095L\_COMB.seq.\*  
91: /cgn2\_6/ptodata/1/pna/US095M\_COMB.seq.\*  
92: /cgn2\_6/ptodata/1/pna/US095N\_COMB.seq.\*  
93: /cgn2\_6/ptodata/1/pna/US095O\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	893	100.0	893	42	US-09-215-435-113
2	640	71.7	1734	50	US-09-354-899-7143
3	640	71.7	1734	51	US-09-371-168-6632
4	559	62.6	878	54	US-09-471-275-3565
5	397	44.5	444	52	US-09-399-720-16879
6	392	43.9	6569	87	US-60-172-360-21713
7	321	35.9	1315	42	US-09-215-435-260
8	321	35.9	1315	67	US-60-069-957-93
9	309	34.6	463	45	US-09-287-618-9033
10	249	27.9	397	51	US-09-371-168-1500
11	231	25.9	418	50	US-09-359-067-47954
12	202	22.6	400	45	US-09-287-618-2702
13	192	21.5	279	37	US-09-070-694-5511
					Sequence 113, App
					Sequence 7143, App
					Sequence 6632, App
					Sequence 3565, App
					Sequence 16879, A
					Sequence 21713, A
					Sequence 260, App
					Sequence 93, Appl
					Sequence 9033, App
					Sequence 47954, A
					Sequence 2702, App
					Sequence 5511, App



```
US-09-354-899-7143
; Sequence 7143, Application US/09354899
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, M. Alexandra
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM A
; FILE REFERENCE: MLN98-35PA
; CURRENT APPLICATION NUMBER: US/09/354,899
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: 60/092,406
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 8631
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7143
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: (1)...(1734)
; OTHER INFORMATION: n = A,T,C or G
US-09-354-899-7143

Query Match          71.7%; Score 640; DB 50; Length 1734;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 690; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 110 tcattgacaggctaaactttgacaaactttgagacaggaatgcatcttgcacgggca 169
Db 259 tcattgacaggctaaactttgacaaactttgagacaggaatgcatcttgcacgggca 318
QY 170 ttgcaaggtacattgagcaggctacgtccactccagcatgaatgagatgctggaggag 229
Db 319 ttgcaaggtacattgagcaggctacgtccactccagcatgaatgagatgctggaggag 378
QY 230 gacatgagatgcggtcatgctgtacacctggcgagctgttcccgccattccccagg 289
Db 379 gacatgagatgcggtcatgctgtacacctggcgagctgttcccgccattccccagg 438
QY 290 tgaatgcaacgagcagcccaaccagtagagatctatgagaagacagtagagtgctgg 349
Db 439 tgaatgcaacgagcagcccaaccagtagagatctatgagaagacagtagagtgctgg 498
QY 470 cctacctctgaccttggcaagttcatcaacatgtttgtctctgcatgagctaaaga 529
Db 619 cctacctctgaccttggcaagttcatcaacatgtttgtctctgcatgagctaaaga 678
QY 530 acataaagtcagcgtcagaatgaccactccgctacaaagggcagcacagttcctgc 589
Db 679 acataaagtcagcgtcagaatgaccactccgctacaaagggcagcacagttcctgc 738
QY 590 ggaagatggcagatccccagtgctccaccagcaactggaagtatccccaggtatgagg 709
Db 799 accaacaagatccccagtgctccaccagcaactggaagtatccccaggtatgagg 858
QY 710 agtctggtgacattgtcaacatctgtgtgattactacagacaagatgtacctga 769
Db 859 agtctggtgacattgtcaacatctgtgtgattactacagacaagatgtacctga 918
QY 770 ctccaggtagaacaatatgtcctccaaggt 800
```

```
Db 919 ctccaggtagaacaatatgtcctccaaggt 949

RESULT 3
US-09-371-168-6632
; Sequence 6632, Application US/093711168
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, M. Alexandra
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM A
; FILE REFERENCE: MLN98-37PA
; CURRENT APPLICATION NUMBER: US/09/371,168
; CURRENT FILING DATE: 1999-08-10
; EARLIER APPLICATION NUMBER: 60/095,907
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/103,145
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 8285
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6632
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: (1)...(1734)
; OTHER INFORMATION: n = A,T,C or G
US-09-371-168-6632
```

```
Query Match          71.7%; Score 640; DB 51; Length 1734;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 690; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 110 tcattgacaggctaaactttgacaaactttgagacaggaatgcatcttgcacgggca 169
Db 259 tcattgacaggctaaactttgacaaactttgagacaggaatgcatcttgcacgggca 318
QY 170 ttgcaaggtacattgagcaggctacgtccactccagcatgaatgagatgctggaggag 229
Db 319 ttgcaaggtacattgagcaggctacgtccactccagcatgaatgagatgctggaggag 378
QY 230 gacatgagatgcggtcatgctgtacacctggcgagctgttcccgccattccccagg 289
Db 379 gacatgagatgcggtcatgctgtacacctggcgagctgttcccgccattccccagg 438
QY 290 tgaatgcaacgagcagcccaaccagtagagatctatgagaagacagtagagtgctgg 349
Db 439 tgaatgcaacgagcagcccaaccagtagagatctatgagaagacagtagagtgctgg 498
QY 350 agccggaggtcaccaagctcatgaagttcatgtattttcagcgcaaggccatcgagcgt 409
Db 499 agccggaggtcaccaagctcatgaagttcatgtattttcagcgcaaggccatcgagcgt 558
QY 410 ttgcagcgaggtgaagcggtgtgccatgcccagcgagcgaggaagactttgtctctgag 469
Db 559 ttgcagcgaggtgaagcggtgtgccatgcccagcgagcgaggaagactttgtctctgag 618
QY 470 cctacctctgaccttggcaagttcatcaacatgtttgtctctgcatgagctaaaga 529
Db 619 cctacctctgaccttggcaagttcatcaacatgtttgtctctgcatgagctaaaga 678
QY 530 acataaagtcagcgtcagaatgaccactccgctacaaagggcagcacagttcctgc 589
Db 679 acataaagtcagcgtcagaatgaccactccgctacaaagggcagcacagttcctgc 738
QY 590 ggaagatggcagatccccagtgctccaccagcaacttccatgttccctgggcca 649
Db 739 ggaagatggcagatccccagtgctccaccagcaacttccatgttccctgggcca 798
QY 650 accaacaagatccccagtgctccaccagcaacttccatgttccctgggcca 709
```

Db	799	acacacacaggaatcaccagtgctccaccagcaacttgaagtgaatcccaagctatgagg	858
Qy	710	agctgctggctgacattgtccaacatctgtgtgattactacgagaacaagatgtacctga	769
Db	859	agctgctggctgacattgtccaacatctgtgtgattactacgagaacaagatgtacctga	918
Qy	770	ctccagctggaacaatatgctcctcaaggt	800
Db	919	ctcccaatgagaacaatatgctcctcaaggt	949

## RESULT

```

US-09-471-275-3565
:
: GENERAL INFORMATION:
: APPLICANT: Hyseq, Inc.
: TITLE OF INVENTION: Novel Contigs Obtained
: TITLE OF INVENTION: From Various Libraries
: FILE REFERENCE: 782
: CURRENT APPLICATION NUMBER: US/09/471,275
: CURRENT FILING DATE: 1999-12-23
: EARLIER APPLICATION NUMBER: US 09/235,076
: EARLIER FILING DATE: 1999-01-20
: EARLIER APPLICATION NUMBER: US 09/234,611
: EARLIER FILING DATE: 1999-01-22
: EARLIER APPLICATION NUMBER: US 09/240,371
: EARLIER FILING DATE: 1999-01-29
: EARLIER APPLICATION NUMBER: US 09/277,227
: EARLIER FILING DATE: 1999-03-25
: EARLIER APPLICATION NUMBER: US 09/271,490
: EARLIER FILING DATE: 1999-03-18
: EARLIER APPLICATION NUMBER: US 09/293,972
: EARLIER FILING DATE: 1999-04-15
: EARLIER APPLICATION NUMBER: US 09/274,861
: EARLIER FILING DATE: 1999-03-23
: EARLIER APPLICATION NUMBER: US 60/125,453
: EARLIER FILING DATE: 1999-03-19
: EARLIER APPLICATION NUMBER: US 60/126,605
: EARLIER FILING DATE: 1999-03-26
: EARLIER APPLICATION NUMBER: US 09/306,350
: EARLIER FILING DATE: 1999-05-07
: EARLIER APPLICATION NUMBER: US 09/399,720
: EARLIER FILING DATE: 1999-09-21
: EARLIER APPLICATION NUMBER: US 09/404,284
: EARLIER FILING DATE: 1999-09-21
: EARLIER APPLICATION NUMBER:
: EARLIER FILING DATE: 1999-12-16
: NUMBER OF SEQ ID NOS: 10451
: SOFTWARE: pt_ct_genes version 1.0
: SEQ ID NO 3565
: LENGTH: 878
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (79)...(747)
: OTHER INFORMATION: similar to g15616320 i
: OTHER INFORMATION: Run with FASTX 3.3+00
US-09-471-275-3565

```

Query Match	62.6%	Score 559;	DB 54;	Length 878;
Best Local Similarity	99.7%	Pred. NO.	1.2e-260;	
Matches 659;	Conservative	0;	Mismatches 2;	Indels 0;
Gaps 0;				

[illegible]

Qy	129	tgacacaaactttgagacacaggaaatcatttgcacgggcattgcgaagtacattgacgca	189
Db	204	tgacacaaactttgagacacaggaaatcatttgcacgggcattgcgaagtacattgacgca	263
Qy	189	ggctacagtcacactccagcatgaatgagatgctggaggaaggacatgagtatgcggtcat	248
Db	264	ggctacagtcacactccagcatgaatgagatgctggaggaaggacatgagtatgcggtcat	323
Qy	249	gctgtacacctggcgcagctgttcccgggccattccccagggtgaaatgcaacgagcagcc	308
Db	324	gctgtacacctggcgcagctgttcccgggccattccccagggtgaaatgcaacgagcagcc	383
Qy	309	caaccgagtagagatctatgagaagacagtagaggtgctggagccggagggtcaccaagct	368
Db	384	caaccgagtagagatctatgagaagacagtagaggtgctggagccggagggtcaccaagct	443
Qy	369	catgaagtcatgtattttcagcgcgaagcccatcgagcgttctgcagcgaggtgaagcg	428
Db	444	catgaagtcatgtattttcagcgcgaagcccatcgagcgttctgcagcgaggtgaagcg	503
Qy	429	gctgtgcctatgccagcgcgaggaagacattgtctctgagcgcctacctctgacccctgg	488
Db	504	gctgtgcctatgccagcgcgaggaagacattgtctctgagcgcctacctctgacccctgg	563
Qy	489	caagttcatcaacatgttttctgctcctggatgagctaaagaacatgaagtgcagcgtcaa	548
Db	564	caagttcatcaacatgttttctgctcctggatgagctaaagaacatgaagtgcagcgtcaa	623
Qy	549	gaatgaccactcgcctacaaagaggcgcacacagttcctctcggaagaatggcagatcccca	608
Db	624	gaatgaccactcgcctacaaagaggcgcacacagttcctctcggaagaatggcagatcccca	683
Qy	609	gtctatccaggagtcgcagaaacttccatgttctcctggcccaaccacacagatcaccca	668
Db	684	gtctatccaggagtcgcagaaacttccatgttctcctggcccaaccacacagatcaccca	743
Qy	669	g 669	
Db	744	g 744	
RESULT 5.			
US-09-399-720-16879			
; Sequence 16879, Application US/09399720			
; GENERAL INFORMATION:			
; APPLICANT: Hyseq, Inc.			
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained			
; TITLE OF INVENTION: From Various Libraries			
; FILE REFERENCE: 20411-777			
; CURRENT APPLICATION NUMBER: US/09/399,720			
; CURRENT FILING DATE: 1999-09-21			
; NUMBER OF SEQ ID NOS: 20869			
; SOFTWARE: Hy-patent.pl Version 3.1			
; SEQ ID NO 16879			
; LENGTH: 444			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; US-09-399-720-16879			

**PRELIM.** 5

```

> US-09-399-720-16879
> Sequence 16879, Application US/09399720
> GENERAL INFORMATION:
> APPLICANT: Hyseq, Inc.
> TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
> FROM VARIOUS LIBRARIES
> FILE REFERENCE: 20411-777
> CURRENT APPLICATION NUMBER: US/09/399,720
> CURRENT FILING DATE: 1999-09-21
> NUMBER OF SEQ ID NOS: 20869
> SOFTWARE: Hy-patent.pl Version 3.1
> SEQ ID NO 16879
> LENGTH: 444
> TYPE: DNA
> ORGANISM: Homo sapiens
> US-09-399-720-16879

```

Query Match 44.5%; Score 397; DB 52; Length 444;  
Best Local Similarity 100.0%; Pred. No. 4.8e-182;  
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	90	cgagcctccaccctctctccatcatgtaccaggctaaactttgacacaaatttggaggacag	149
Db	48	cgagcctccaccctctctccatcatgtaccaggctaaactttgacacaaatttggaggacag	107
Qy	150	gaatgcatttgtcacgggcatttgcgaaggctacatttgacgagcgttcacagtccaactccagcat	209
Db	108	gaatgcatttgtcacgggcatttgcgaaggctacatttgacgagcgttcacagtccaactccagcat	167
Qy	210	gaatggaatctctgagaaagacacatdaagtatgcggctcatctctctacacttgcggaagctc	269

```
Db 168 gaatgagatgctggggaaggacatgagatgctggtctatgctgctacacctggcgagctg 227
QY 270 ttcccgggccattccccaggtgaaatgcaacgagcagcagcccaaccaggtagagatctatga 329
Db 228 ttcccgggccattccccaggtgaaatgcaacgagcagcagcccaaccaggtagagatctatga 287
QY 330 gaagacagtagaggtgctggagccggagggtcaccgaagctcatgaagttcatattttca 389
Db 288 gaagacagtagaggtgctggagccggagggtcaccgaagctcatgaagttcatattttca 347
QY 390 gcgcaggcccatcgagcgggttctgcagcgaggtgaagcggtgtgcctagccgagcgag 449
Db 348 gcgcaggcccatcgagcgggttctgcagcgaggtgaagcggtgtgcctagccgagcgag 407
QY 450 gaagacattgtctctgagcctactctctgacctt 486
Db 408 gaaggacttgtctctgagcctactctctgacctt 444

RESULT 6
US-60-172-360-21713
; Sequence 21713, Application US/60172360
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Dlep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polymorph
; FILE REFERENCE: GX-0007 P
; CURRENT APPLICATION NUMBER: US/60/172,360
; CURRENT FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 29838
; SOFTWARE: PERL Program
; SEQ ID NO 21713
; LENGTH: 6569
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 232778.5
; NAME/KEY: unsure
; LOCATION: 390, 4145-4174, 4903
; OTHER INFORMATION: a, t, c, g, or other
US-60-172-360-21713
```

```
Query Match 43.9%; Score 392; DB 87; Length 6569;
Best Local Similarity 99.8%; Pred. No. 1.5e-179;
Matches 442; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 358 gtccaccaagctcatgagttcatgtatttttcagcgcgaagcccatcgagcggttctgcagc 417
Db 480 gtccaccaagctcatgagttcatgtatttttcagcgcgaagcccatcgagcggttctgcagc 539
QY 418 gagggtgaagcggctgtgcctatgccagcgcgaggaagatttctctgagcctacctc 477
Db 540 gagggtgaagcggctgtgcctatgccagcgcgaggaagatttctctgagcctacctc 599
QY 478 ctgacccttggcaagttcatcaatgtttgtctgctgagatgagctaaagaacatgaag 537
Db 600 ctgacccttggcaagttcatcaatgtttgtctgctgagatgagctaaagaacatgaag 659
QY 538 tgcagcgtcaagaatgaccactccgctacaaagagggcagcacagttctctgggaagatg 597
Db 660 tgcagcgtcaagaatgaccactccgctacaaagagggcagcacagttctctgggaagatg 719
QY 598 gcagatccccagttctatccagagtgcgagaacaccttccatgttctctggccaaccacaac 657
Db 720 gcagatccccagttctatccagagtgcgagaacaccttccatgttctctggccaaccacaac 779
QY 658 aggateaccacagtgctctccaccagcaacttgaagtgtatccagggtatgaggagtgctg 717
```

```
Db 780 aggaacccccagtgctccaccagcaacttgaagtgtatcccaaggctatgaggagtgctg 839
QY 718 gctgacattgtcaacatctgtgtgattactacgagaacaagatgtacctgactccaggt 777
Db 840 gctgacattgtcaacatctgtgtgattactacgagaacaagatgtacctgactccaggt 899
QY 778 gagaacaatatgctctctcaaggt 800
Db 900 gagaacaatatgctctctcaaggt 922

RESULT 7
US-09-215-435-260
; Sequence 260, Application US/09215435A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric,
; APPLICANT: Bouqueleret Lydie
; TITLE OF INVENTION: Extended CDNAS
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215,435A
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/069,957
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 260
; LENGTH: 1315
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 123..302
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 123..176
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 4.30000019073486
; OTHER INFORMATION: seq WTCLKSFSPSTSS/HA
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1279..1284
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 1301..1312
US-09-215-435-260
```

```
Query Match 35.9%; Score 321; DB 42; Length 1315;
Best Local Similarity 100.0%; Pred. No. 4e-145;
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 562 gctacaagaggcgacagttctctgcggaagatggcagatccccagttctatccaggag 621
Db 656 gctacaagaggcgacagttctctgcggaagatggcagatccccagttctatccaggag 715
QY 622 tcgcagaacctttccatgttctctggcccaaccacacagatcacccagttgtctccaccag 681
Db 716 tcgcagaacctttccatgttctctggcccaaccacacagatcacccagttgtctccaccag 775
QY 682 caacttgaagtatccccagggtatgaggagctgctggctgacattgtcaacatctgtgtg 741
Db 776 caacttgaagtatccccagggtatgaggagctgctggctgacattgtcaacatctgtgtg 835
QY 742 gattactacgagaacaagatgtacctgactccagtgagaacaacatgtctctcaaggta 801
```

Db 836 gattactacgagacaagatgtacctgactcccagtgagaaaacatatgtctctcaaggta 895  
 QY 802 aaactccctgagcgcacccatgagcctggcttacccctcaccctctctcttatta 861  
 Db 896 aaactccctgagcgcacccatgagcctggcttacccctcaccctctctcttatta 955  
 QY 862 aaatccggttttaaaaaaa 882  
 Db 956 aaatccggttttaaaaaaa 976

RESULT 8

US-60-069-957-93  
 ; Sequence 93, Application US/60069957  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edwards, Jean-Baptiste Dumas Milne  
 ; APPLICANT: Duclert, Aymeric  
 ; TITLE OF INVENTION: EXTENDED CDNAS FOR SECRETED PROTEINS  
 ; NUMBER OF SEQUENCES: 381  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear  
 ; STREET: 501 West Broadway  
 ; CITY: San Diego  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 92101-3505  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy Disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: Win95  
 ; SOFTWARE: Word  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/60/069,957  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Israel, Ned A.  
 ; REGISTRATION NUMBER: 29,655  
 ; REFERENCE/DOCKET NUMBER: GENSET.019PR  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (619) 235-8550  
 ; TELEFAX: (619) 235-0176  
 ; INFORMATION FOR SEQ ID NO: 93:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1315 base pairs  
 ; TYPE: NUCLEIC ACID  
 ; STRANDEDNESS: DOUBLE  
 ; TOPOLOGY: LINEAR  
 ; MOLECULE TYPE: CDNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Homo Sapiens  
 ; TISSUE TYPE: Brain  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 123..302  
 ; FEATURE:  
 ; NAME/KEY: sig\_peptide  
 ; LOCATION: 123..176  
 ; IDENTIFICATION METHOD: Von Heijne matrix  
 ; OTHER INFORMATION: score 4.3  
 ; OTHER INFORMATION: seq WTCLKSFPSPSS/HA  
 ; FEATURE:  
 ; NAME/KEY: poly\_a\_signal  
 ; LOCATION: 1279..1284  
 ; FEATURE:  
 ; NAME/KEY: poly\_a  
 ; LOCATION: 1301..1312  
 ; IDENTIFICATION METHOD: blastn  
 ; FEATURE:  
 ; NAME/KEY: est  
 ; LOCATION: 628..868  
 ; IDENTIFICATION METHOD: blastn  
 ; OTHER INFORMATION: identity 95

; ; OTHER INFORMATION: region 1..241  
 ; ; OTHER INFORMATION: id AAL43673  
 US-60-069-957-93  
 Query Match 35.9%; Score 321; DB 67; Length 1315;  
 Best Local Similarity 100.0%; Pred. No. 4e-145;  
 Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 562 gctacaagagggcagcacagttctgcggaagatggcagatccccagctctccagag 621  
 Db 656 GCTACAAGAGGGCAGCACAGTTCTTCGCGAAGATGGCAGATCCCCAGTCTATCCAGGAG 715  
 QY 622 tcgcagaacctttccatgttctctggccaaaccacacacagcagatcaccagtgctccaccag 681  
 Db 716 TCGCAGAACCTTTCCATGTCTCTGGCCAAACCAACAGGATCACCAGTGTCTCCACCAG 775  
 QY 682 caacttgaagtatccccaggtatgaggagctgctggctgacattgtcaaatctctgtg 741  
 Db 776 CAAC TTGAAGTATCCAGGCTATGAGGAGCTGCTGGCTGACATTGTCAACATCTGTGTG 835  
 QY 742 gattactacgagaacaagatgtacctgactccagtgagaaacatatgtctcctcaaggta 801  
 Db 836 GATTACTACGAGACAAGATGTACCTGACTCCAGTCCAGTGAACATATGCTCTCAGGTA 895  
 QY 802 aaactccctgagcgcacccatgagcctggcttacccctcaccctctctcttatta 861  
 Db 896 AAAC TCCCTGAGCGGCACCCATGGAGCTGGGCTTACCTCTCACCTTCTTCTATTATTA 955  
 QY 862 aaatccggttttaaaaaaa 882  
 Db 956 AAATCCGTTTAAAAAACAA 976

RESULT 9

US-09-287-618-9033  
 ; Sequence 9033, Application US/09287618  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hyseq, Inc.  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
 ; FILE REFERENCE: 20411-768  
 ; CURRENT APPLICATION NUMBER: US/09/287,618  
 ; CURRENT FILING DATE: 1999-04-02  
 ; NUMBER OF SEQ ID NOS: 35865  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 9033  
 ; LENGTH: 463  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(463)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-287-618-9033  
 Query Match 34.6%; Score 309; DB 45; Length 463;  
 Best Local Similarity 99.7%; Pred. No. 2.5e-139;  
 Matches 359; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 9 cagcagctcacccctggaagatgccctgtccaaactgacccctgcttgaagagcttccct 68  
 Db 104 cagcagctcacccctggaagatgccctgtccaaactgacccctgcttgaagagcttccct 163  
 QY 69 cccgcaccagcagccatgcacgcctccacccctccctccatcatgtaccaggttaact 128  
 Db 164 cccgcaccagcagccatgcacgcctccacccctccctccatcatgtaccaggttaact 223  
 QY 129 tgacacaaactttgaggacaggaatgcatgttcacgggcatgtgcaaggtacattgagca 188  
 Db 224 tgacacaaactttgaggacaggaatgcatgttcacgggcatgtgcaaggtacattgagca 283



Qy	189	gggtacacgtccatccagcatgaatgagatgctggagggaagacatgagatgcggtcat	248
Db	284	gggtacagtcctaccagcatgaatgagatgctggagggaagacatgagatgcggtcat	343
Qy	249	gctgtacacctgctgcagctgttcccggggccattccccagggtgaaatgcaacgagagcgc	308
Db	344	gcgtcacctggcgcagctgttcccggggccattccccagggtgaaatgcaacgagagcgc	403
Qy	309	caaccagtgtagatctatgagaagacagtagagggtctctggaccgcgaggtcaccaagct	368
Db	404	caaccagtgtagatctatgagaagacagtagagggtctctggaccgcgaggtcaccaagct	463

```

RESULT 10
US-09-371-168-1500
; Sequence 1500, Application US/09371168
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, M. Alexandra
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM A
; FILE REFERENCE: MLN98-370A
; CURRENT APPLICATION NUMBER: US/09/371.168
; CURRENT FILING DATE: 1999-08-10
; EARLIER APPLICATION NUMBER: 60/095,907
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/103,145
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 8285
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1500
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-371-168-1500

```

	Query Match	27.98;	Score 249;	DB 51;	Length 397;
	Best Local Similarity	99.7%;	Pred. No. 3.2e-110;		
	Matches 299;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	501	catgtttgctgtcctgtagtactgaagaacatgaagtgcagcgtccaagaatgaccactc	560		
Db	4	catgtttgctgtcctgtagtactgaagaacatgaagtgcagcgtccaagaatgaccactc	63		
QY	561	cgctacaagaaggcagcacacagttcctgcggaagatggcagatccccagttatccaggga	620		
Db	64	tgctacaagaaggcagcacagttcctgcggaagatggcagatccccagttatccaggga	123		
QY	621	gtcgcagaacctttccatgtttccttggccaaaccacaacaggatcacccagtggtctccacca	680		
Db	124	gtcgcagaacctttccatgtttccttggccaaaccacaacaggatcacccagtggtctccacca	183		
QY	681	gcaactggaagtgtcccagctatgaggagctgctggctgacattgtccaacatctgtgt	740		
Db	184	gcaactggaagtgtcccagctatgaggagctgctggctgacattgtccaacatctgtgt	243		
QY	741	ggattactacgagaacaagatgtacctgactccccagtgagaacaatatgtcctccaaggt	800		
Db	244	ggattactacgagaacaagatgtacctgactccccagtgagaacaatatgtcctccaaggt	303		

RESULT 11  
US-09-359-067-47954.  
; Sequence 47954, Application US/09359067  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-747COM1  
; CURRENT APPLICATION NUMBER: US/09/359,067  
; CURRENT FILING DATE: 1999-07-22  
; EARLIER APPLICATION NUMBER: US 09/131,598  
; EARLIER FILING DATE: 1998-08-10

```

; NUMBER OF SEQ ID NOS: 49786
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 47954
; LENGTH: 418
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(418)
; OTHER INFORMATION: n = A,T,C or G
US-09-359-067-47954

```

Query Match	25.9%	Score 231;	DB 50;	Length 418;
Best Local Similarity	99.6%;	Pred. No. 1.8e-101;		
Matches 281;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY 388	cagcgcaaggccatcgagcggttctgcagcgaggtgaagcggctgtgcatcgcgagcgc	447		
Db 3	cagcgcaaggccatcgagcggttctgcagcgaggtgaagcggctgtgcatcgcgagcgc	62		
QY 448	aggagagcattgtctctgagcgctacctcctcagcccttggcaagttcatcaacatgttt	507		
Db 63	aggagagcattgtctctgagcgctacctcctcagcccttggcaagttcatcaacatgttt	122		
QY 508	gctctctggatgagctaaagaaacatgaagtgcagcgtcaagaatgaccactcgcgctac	567		
Db 123	gctctctggatgagctaaagaaacatgaagtgcagcgtcaagaatgaccactcgcgctac	182		
QY 568	aagagggcgagcacagtttctctcggaagatggcagatccccagttctatccaggagtcgag	627		
Db 183	aagagggcgagcacagtttctctcggaagatggcagatccccagttctatccaggagtcgag	242		
QY 628	aacctttccatgttctctcggccaacccaacaaaggatcacccag	669		
Db 243	aacctttccatgttctctcggccaacccaacaaaggatcacccag	284		

```

RESULT      12
US-09-287-618-2702
; Sequence 2702, Application US/09287618
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-768
; CURRENT APPLICATION NUMBER: US/09/287,618
; CURRENT FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 35865
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2702
; LENGTH: 400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-287-618-2702

```

	Query Match	22.6%	Score 202;	DB 45;	Length 400;
	Best Local Similarity	100.0%;	Prod. No. 2.1e-87;		
	Matches 202;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	9	cacgcagtcaccctggaagatgccctgtccacagtcggacctgtgaagagcttccct	68		
Db	81	cacgcagtcaccctggaagatgccctgtccacagtcggacctgtgaagagcttccct	140		
Qy	69	cccgaccagcagccatgcatacgcagctccaccttccctcatgtaccagggttaactt	128		
Db	141	cccgaccagcagccatgcatacgcagctccaccttccctcatgtaccagggttaactt	200		
Qy	129	tgacacaaactttgagacaggaatgatttgtcacggggcaattcacagggtacattgagca	188		
Db	201	tgacacaaactttgagacaggaatgatttgtcacggggcaattcacagggtacattgagca	260		

QY 189 ggctacagtcactccagcatg 210  
|||||  
Db 261 ggctacagtcactccagcatg 282

## RESULT 13

US-09-070-694-5511  
; Sequence 5511, Application US/09070694  
; GENERAL INFORMATION:  
; APPLICANT: Gooding, Douglas H.  
; APPLICANT: Stuve, Laura L.  
; APPLICANT: Stuart, Susan G.  
; APPLICANT: Ito, Laura Y.  
; APPLICANT: Akerblom, Ingrid E.  
; APPLICANT: Delegeane, Angelo M.  
; APPLICANT: Naughton, Rebecca E.  
; APPLICANT: Klingler, Tod M.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM  
; TITLE OF INVENTION: TH2 CELLS  
; NUMBER OF SEQUENCES: 5899  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/070,694  
; FILING DATE: HERewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/044,030  
; FILING DATE: MAY 2, 1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CERRONE, MICHAEL C.  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PD-0364P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 5511:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 279 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; IMMEDIATE SOURCE:  
; CLONE: 3003476H1  
US-09-070-694-5511

Query Match 21.5%; Score 192; DB 37; Length 279;  
Best Local Similarity 100.0%; Pred. No. 1.5e-82;  
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 cagcagcgtccactggagatccctgtccaaagctggacctgttgagagagcttccct 68  
|||||  
Db 49 CAGCGACGTCACCCCTGGAGATGCCCTGTCCAAAGCTGGACCTCTTGAAGAGCTTCCCT 108  
QY 69 ccccgaccagcagccatgcatgcagcctccacaccttccatcatgtaccaggttaact 128  
|||||  
Db 109 CCCCAGCAGCAGCCATGATCGAGCTCCACCTTCTCCATCATGTACCAGGCTAACTT 168  
QY 129 tgacacaaactttgagacaggaatgcatttgcacgggcatgtcgaaggtacattgagca 188  
|||||

Db 169 TGACACAAACTTTGAGGACAGGAATGCATTTGTACGGGCGCATTCGAAGGTACATTGAGCA 228  
QY 189 ggctacagtcacca 200  
|||||  
Db 229 GGCTACAGTCCA 240

## RESULT 14

US-60-044-030-3917  
; Sequence 3917, Application US/60044030  
; GENERAL INFORMATION:  
; APPLICANT: Gooding, Douglas H.  
; APPLICANT: Stuve, Laura L.  
; APPLICANT: Stuart, Susan G.  
; APPLICANT: Ito, Laura Y.  
; APPLICANT: Akerblom, Ingrid E.  
; APPLICANT: Delegeane, Angelo M.  
; APPLICANT: Naughton, Rebecca E.  
; APPLICANT: Klingler, Tod M.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM  
; TITLE OF INVENTION: TH2 CELLS  
; NUMBER OF SEQUENCES: 4305  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/60/044,030  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CERRONE, MICHAEL C.  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PD-0364P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 855-0555  
; TELEFAX: (415) 845-4166  
; INFORMATION FOR SEQ ID NO: 3917:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 279 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; IMMEDIATE SOURCE:  
; CLONE: 3003476H1  
US-60-044-030-3917

Query Match 21.5%; Score 192; DB 64; Length 279;  
Best Local Similarity 100.0%; Pred. No. 1.5e-82;  
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 cagcagcgtccactggagatccctgtccaaagctggacctgttgagagagcttccct 68  
|||||  
Db 49 CAGCGACGTCACCCCTGGAGATGCCCTGTCCAAAGCTGGACCTCTTGAAGAGCTTCCCT 108  
QY 69 ccccgaccagcagccatgcatgcagcctccacaccttccatcatgtaccaggttaact 128  
|||||  
Db 109 CCCCAGCAGCAGCCATGATCGAGCTCCACCTTCTCCATCATGTACCAGGCTAACTT 168  
QY 129 tgacacaaactttgagacaggaatgcatttgcacgggcatgtcgaaggtacattgagca 188  
|||||  
Db 169 TGACACAAACTTTGAGGACAGGAATGCATTTGTACGGGCGCATTCGAAGGTACATTGAGCA 228



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 14:44:27 ; Search time 3022.95 Seconds  
(without alignments)  
1115.357 Million cell updates/sec

Title: US-09-215-435-113  
Perfect score: 893  
Sequence: 1 gccatgacgcgcgcgcac.....aaaaaaaaaaaaaaaaaaaa 893

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4538634 seqs, 1887831982 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*

44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_gss1:\*  
80: gb\_gss2:\*  
81: gb\_gss3:\*  
82: gb\_gss4:\*  
83: em\_gss1:\*  
84: em\_gss2:\*  
85: em\_gss3:\*  
86: em\_gss4:\*  
87: gb\_gss5:\*  
88: gb\_gss6:\*  
89: gb\_gss7:\*  
90: gb\_gss8:\*  
91: gb\_gss9:\*  
92: em\_gss5:\*  
93: em\_gss6:\*  
94: em\_gss7:\*  
95: em\_gss8:\*  
96: em\_gss9:\*  
97: em\_gss10:\*  
98: em\_gss11:\*  
99: gb\_gss10:\*  
100: gb\_gss11:\*  
101: em\_gss12:\*  
102: gb\_gss12:\*  
103: gb\_gss13:\*  
104: gb\_gss14:\*  
105: gb\_gss15:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	256	28.7	472	62	AI929491
					AI929491 au60f02.y



REFERENCE 1 (bases 1 to 506)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3187906.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/dbbrp/image/image.html](http://www-bio.llnl.gov/dbbrp/image/image.html)

Insert Length: 876 Std Error: 0.00  
 Seq primer: ~40UP from Gibco  
 High quality sequence stop: 459.

#### FEATURES

Location/Qualifiers  
 1..506  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2137854"  
 /clone\_lib="NCI-CGAP\_K1d11"  
 /lab\_host="DH10B"  
 /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with  
 a modified polylinker; Site:1: Not 1; Site:2: Eco RI;  
 Plasmid DNA from the normalized library NCI-CGAP\_K1d3 was  
 prepared, and ss circles were made in vitro. Following HAP  
 purification, this DNA was used as tracer in a subtractive  
 hybridization reaction. The driver was PCR-amplified cDNAs  
 from a pool of 5,000 clones made from the same library  
 (cloneIDs 1322376-1323911, 1456007-1456775, and  
 1500552-1502855). Subtraction by Bento Soares and M.  
 Fatima Bonaldo."  
 123 a 156 c 117 g 110 t

#### BASE COUNT

ORIGIN

Query Match 22.7%; Score 203; DB 47; Length 506;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-86;  
 Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 cagcagctcacctggaagatgccctgtccacgtggacctgcttgaagagcttccct 68  
 Db 161 CAGCGACGTCACCTGGAAGATGCCCTGTCCAAGTGGACCTGTTGAAGAGCTTCCCT 220  
 QY 69 ccccgaccagcagcatgcatgagcctccaccttctccatcatgataccaggctaactt 128  
 Db 221 CCCCGACAGCAGCAGCATGATCGAGCTCCACCTTCTCCATCATGTACCAAGCTAACTT 280  
 QY 129 tgacacaaacttgagcaggaatcattgtcacggcattgcaaggtacattgagca 188  
 Db 281 TGACACAACTTTGAGCAGAGGAATGCATTTGTCAGGGCATGCAAGGTACATTGAGCA 340  
 QY 189 ggcacagtcacattccacagcatga 211  
 Db 341 GGCTACAGTCCACTCCAGCATGA 363

#### RESULT

3  
 LOCUS AI909739/c 186 bp mRNA EST  
 DEFINITION QV-BT223-040599-063 BT223 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION AI909739  
 VERSION AI909739.1 GI:6500419  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

#### REFERENCE

1 (bases 1 to 186)  
 AUTHORS HCOP <http://www.ludwig.org.br/ORESTES>.  
 TITLE The FAPESP/LICR Human Cancer Genome Project  
 JOURNAL Unpublished (1999)  
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3189519.  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/seq/gethtml.pl?tl-QVst2-QV-BT223-063.htmlst3-040599st4-1>)  
 Seq primer: puc 18 forward.

#### FEATURES

Location/Qualifiers  
 1..186  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="BT223"  
 /sex="female"  
 /dev\_stages="Adult"  
 /note="Organ: breast; Vector: puc18; Site:1: SmaI; Site:2:  
 SmaI; A mini-library was made by cloning products derived  
 from ORESTES PCR (U.S. Letters Patent application No.  
 196,716 - Ludwig Institute for Cancer Research) profiles  
 into the puc 18 vector. Reverse transcription of tissue  
 mRNA and cDNA amplification were performed under low  
 stringency conditions."  
 38 a 55 c 47 g 46 t

#### BASE COUNT

ORIGIN

Query Match 16.2%; Score 145; DB 62; Length 186;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-58;  
 Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 416 gcgaggtgaagcggctgtgccatgccagcgcagcaggaagacttctctgagcctacc 475  
 Db 186 GCGAGGTGAAGCGGCTGTGCCATGCCAGCGCAGGAGGACTTTCTCTGAGGCTACC 127  
 QY 476 tcttgaccttggcagtgatcacaatgtttgtctgctgagtgagctaaagaacatga 535  
 Db 126 TCCTGACCTTGGCAAGTTCATCAACATGTTTGTCTGTCTGTGATGAGCTAAAGACATGA 67  
 QY 536 agtcagcgcgtcaagaatgaccactc 560  
 Db 66 AGTGACGCGTCAAGAATGACCACTC 42

#### RESULT

4  
 LOCUS AI834230/c 217 bp mRNA EST  
 DEFINITION RCO-HT0012-030799-001-H04 HT0012 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION AI834230  
 VERSION AI834230.1 GI:5456573  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 217)  
 AUTHORS HCOP <http://www.ludwig.org.br/ORESTES>.  
 TITLE The FAPESP/LICR Human Cancer Genome Project  
 JOURNAL Unpublished (1999)  
 COMMENT On May 18, 1998 this sequence version replaced gi:3136775.  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research





BASE COUNT	188 a	134 c	152 g	140 t	CalTech Human PAC Library D"
ORIGIN					
Query Match	7.3%; Score 65; DB 102; Length 614;				
Best Local Similarity	100.0%; Pred.No. 1.6e-20;				
Matches 65; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
QY	146	acaggaatgattgtcacggcgatgcaagggtacattgacagcagggtacagtcacatccca	205		
Db	614	ACAGGAATGCATTGTCACGGGCAATGTCAGGTACATTGACAGGCTACAGTCCACTCCA	555		
QY	206	gcgatg	210		
Db	554	GCATG	550		
RESULT	8				
AI882137					
LOCUS	AI882137 543 bp mRNA EST 22-JUL-1999				
DEFINITION	ul30h11.v1 Sugano mouse kidney mkia Mus musculus cDNA clone				
ACCESSION	IMAGE:2099877 5' similar to TR:088558 O88558 SHYC. ;, mRNA				
VERSION	sequence.				
KEYWORDS	EST.				
SOURCE	AI882137.1 GI:5567226				
ORGANISM	house mouse.				
REFERENCE	Mus musculus				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	1 (bases 1 to 543)				
JOURNAL	Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.				
COMMENT	The WashU-NCI Mouse EST Project 1999 Unpublished (1999) On Dec 20, 1995 this sequence version replaced gi:1133710. Other_ESTs: ul30h11.x1 Contact: Marra M/WashU-NCI Mouse EST Project 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:996809 Seq primer: custom primer used High quality sequence stop: 503.				
FEATURES	Location/Qualifiers				
source	1. .543 /organism="Mus musculus" /strain="C57BL" /db_xref="taxon:10090" /clone="IMAGE:2099877" /clone_lib="Sugano mouse kidney mkia" /sex="female" /dev_stage="adult" /lab_host="DH10B" /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII (CACTGTGG); Site_2: draIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTCGCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for				

sequencing: 5' end primer CTTCTGCTCTAAAGCTGGC and 3' end primer CGACCTGCGAGCTGACACA."  
 BASE COUNT 121 a 167 c 161 g 94 t

Query Match 5.6%; Score 50; DB 62; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-13;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 142 gagacaggaatgcatgttcacggcgattgcaagggtacattgagcaggc 191  
 Db 229 GAGACAGGAATGATTTGTCACGGCGATTGCAAGGTACATTGAGCAGGC 278

RESULT 9  
 AI006658 721 bp mRNA EST 12-JUN-1998  
 LOCUS uel6a08.y1 Sugano mouse embryo mewa Mus musculus cDNA clone  
 DEFINITION IMAGE:1480502 5' similar to TR:Q14467 Q14467 KIAA0068 ; , mRNA  
 sequence.

ACCESSION AI006658.1 GI:3216267  
 VERSION AI006658  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 721)  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.

TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On May 8, 1995 this sequence version replaced gi:800369.  
 Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LML; contact the  
 IMAGE Consortium (info@image.llni.gov) for further information.  
 MGI:928858

Possible reversed clone: similarity on wrong strand  
 Seq primer: primer name ambiguous  
 High quality sequence stop: 465.

## FEATURES

Location/Qualifiers  
 1..721  
 /organism="Mus musculus"  
 /strain="C57BL"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1480502"  
 /clone\_lib="Sugano mouse embryo mewa"  
 /dev\_stage="embryo, 14 dpc"  
 /lab\_host="DH10B"

/note="Vector: pME18S-FL3; Site\_1: DraIII (CACTGTGTG); Site\_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTT] double-stranded cDNA was ligated to a DraIII adaptor [ATGTGGCCTTTT] and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAGCTGGC and 3' end primer CGACCTGCGAGCTGACACA."

BASE COUNT 171 a 212 c 203 g 135 t

## ORIGIN

Query Match 5.6%; Score 50; DB 41; Length 721;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-13;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 142 gagacaggaatgcatgttcacggcgattgcaagggtacattgagcaggc 191  
 Db 227 GAGACAGGAATGATTTGTCACGGCGATTGCAAGGTACATTGAGCAGGC 276

RESULT 10  
 AU079605 851 bp mRNA EST 20-OCT-1999  
 LOCUS AU079605 Sugano mouse brain mnchb Mus musculus cDNA clone MNCb-4535  
 DEFINITION 5', mRNA sequence.

ACCESSION AU079605  
 VERSION AU079605.1 GI:6084360  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 851)

AUTHORS Hashimoto, K., Kusuda, J., Toyoda, A., Tanuma, R., Ito, A., Hirata, M.,  
 Suzuki, Y., Sasaki, M. and Sugano, S.  
 TITLE Isolation of full-length cDNA clones from a mouse brain cDNA  
 library made by oligo-capping method  
 JOURNAL Unpublished (1999)  
 COMMENT On Feb 22, 1999 this sequence version replaced gi:4282828.  
 Contact: Katsuyuki Hashimoto  
 Division of Genetic Resources  
 National Institute of Infectious Diseases  
 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan  
 Email: khashi@nih.go.jp  
 URL: http://www.nih.go.jp/yoken/genbank/.

LOCATION/Qualifiers  
 1..851  
 /organism="Mus musculus"  
 /strain="C57BL"  
 /db\_xref="taxon:10090"  
 /clone="MNCb-4535"  
 /clone\_lib="Sugano mouse brain mnchb"  
 /sex="female"  
 /dev\_stage="adult"  
 /lab\_host="TOP10"

## FEATURES

/note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTT] double-stranded cDNA was ligated to a DraIII adaptor [ATGTGGCCTTTT] and cloned into distinct DraIII sites of the pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5 kb. Library was constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer for sequencing: 5' end primer [CTTCTGCTCTAAAGCTGGC]"

BASE COUNT 226 a 215 c 198 g 162 t 50 others

## ORIGIN

Query Match 4.9%; Score 44; DB 69; Length 851;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;  
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 gagacaggaatgcatgttcacggcgattgcaagggtacattgacattga 185  
 Db 238 GAGACAGGAATGATTTGTCACGGCGATTGCAAGGTACATTGA 281

RESULT 11  
 AA143673

```

LOCUS      AA143673      241 bp      mRNA      EST      04-DEC-1996
DEFINITION z063h06.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
IMAGE:591803 5' similar to TR:G559703 G559703 MRNA ;, mRNA
sequence.
ACCESSION  AA143673
VERSION     AA143673.1 GI:1713061
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 241)
AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,N.,
            Maridis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
            Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
            and Marra,M.
            Generation and analysis of 280,000 human expressed sequence tags
            Genome Res. 6 (9), 807-828 (1996)
            9704478
            On Nov 29, 1993 this sequence version replaced gi:634620.
            Contact: Wilton RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28m13 rev2 from Amersham
            High quality sequence stop: 1.
FEATURES   source
            1..241
            /organism="Homo sapiens"
            /db_xref="GBD:4623189"
            /db_xref="taxon:9606"
            /clone="IMAGE:591803"
            /lab_host="SOLR cells (kanamycin resistant)"
            /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
            EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
            oligo dt. Pancreatic adenocarcinoma cell line. Average
            insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
            sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'
            CTCGAGTTTCTTTTCTTTTCTTTT 3'"
BASE COUNT 66 a 67 c 53 g 47 t 8 others
ORIGIN
Query Match 4.7%; Score 42; DB 29; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 tctccaccagcaactgaagtgcagctccagctatgagagct 713
      |||||||
Db 139 TCTCCACCAGCAACTGACGTGATCCCGCAGCTATGAGAGCT 180
      |||||||

RESULT 12
AU035506
LOCUS      AU035506      763 bp      mRNA      EST      08-OCT-1998
DEFINITION AU035506 Sugano mouse brain mncb Mus musculus cDNA clone MNCB-0526,
            mRNA sequence.
ACCESSION  AU035506
VERSION     AU035506.1 GI:3718509
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 763)
AUTHORS   Hata,H., Yamaguchi,R., Tateyama,S. and Sugano,S.
            Construction of mouse full length-enriched cDNA libraries
            Unpublished (1998)
            On Jan 19, 1998 this sequence version replaced gi:2045870.
            Contact: Katsuyuki Hashimoto
            Division of Genetic Resources
            National Institute of Infectious Diseases
            23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
            Email: khashi@nih.go.jp.
            Location/Qualifiers
            1..763
            /organism="Mus musculus"
            /strain="C57BL"
            /db_xref="taxon:10090"
            /clone="MNCB-0526"
            /clone_lib="Sugano mouse brain mncb"
            /sex="female"
            /dev_stage="adult"
            /lab_host="TOP10"
            /note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA
            was primed with an oligo(dT) primer
            [ATGTGGCGCTTTTCTTTTCTTTTCTTTT] double-stranded cDNA was
            ligated to a braill adaptor [TGTGGCGCTTTTCTTTTCTTTTCTTTT] , digested
            and cloned into distinct draiii sites of the pME18S-FL3.
            XhoI sites just outside the DraIII sites can be used to
            isolate the cDNA insert. Size selection was performed to
            exclude fragments <1.5 kb. Library was constructed by
            Sugano et al. (University of Tokyo, Institute of Medical
            Science). Custom primer for sequencing: 5' end primer
            [CTTCTGCTCTAAAGCTGCG]"
BASE COUNT 180 a 166 c 150 g 176 t 91 others
ORIGIN
Query Match 3.6%; Score 32; DB 43; Length 763;
Best Local Similarity 100.0%; Pred. No. 7.4e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 tctccatcatgtaccagcttaacttgacac 134
      |||||||
Db 191 TCTCCATCATGTACCAGGCTTACTTTTGACAC 222
      |||||||

RESULT 13
AA455674
LOCUS      AA455674      91 bp      mRNA      EST      07-APR-1999
DEFINITION aa22a01.r1 NCI-CGAP-CCBI Homo sapiens cDNA clone IMAGE:813960 5'
            similar to TR:G559703 G559703 MRNA ;, mRNA sequence.
ACCESSION  AA455674
VERSION     AA455674.1 GI:2178450
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 91)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            On Nov 29, 1993 this sequence version replaced gi:503228.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Insert Length: 1221 Std Error: 0.00
            Seq primer: -28m13 rev2 ET from Amersham

```

```

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 763)
Sasaki,Z., Suzuki,Y., Watanabe,M., Imai,J., Shibui,A., Yoshida,K.,
Hata,H., Yamaguchi,R., Tateyama,S. and Sugano,S.
Construction of mouse full length-enriched cDNA libraries
Unpublished (1998)
On Jan 19, 1998 this sequence version replaced gi:2045870.
Contact: Katsuyuki Hashimoto
Division of Genetic Resources
National Institute of Infectious Diseases
23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
Email: khashi@nih.go.jp.
            Location/Qualifiers
            1..763
            /organism="Mus musculus"
            /strain="C57BL"
            /db_xref="taxon:10090"
            /clone="MNCB-0526"
            /clone_lib="Sugano mouse brain mncb"
            /sex="female"
            /dev_stage="adult"
            /lab_host="TOP10"
            /note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA
            was primed with an oligo(dT) primer
            [ATGTGGCGCTTTTCTTTTCTTTTCTTTT] double-stranded cDNA was
            ligated to a braill adaptor [TGTGGCGCTTTTCTTTTCTTTTCTTTT] , digested
            and cloned into distinct draiii sites of the pME18S-FL3.
            XhoI sites just outside the DraIII sites can be used to
            isolate the cDNA insert. Size selection was performed to
            exclude fragments <1.5 kb. Library was constructed by
            Sugano et al. (University of Tokyo, Institute of Medical
            Science). Custom primer for sequencing: 5' end primer
            [CTTCTGCTCTAAAGCTGCG]"
BASE COUNT 180 a 166 c 150 g 176 t 91 others
ORIGIN
Query Match 3.6%; Score 32; DB 43; Length 763;
Best Local Similarity 100.0%; Pred. No. 7.4e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 tctccatcatgtaccagcttaacttgacac 134
      |||||||
Db 191 TCTCCATCATGTACCAGGCTTACTTTTGACAC 222
      |||||||

RESULT 13
AA455674
LOCUS      AA455674      91 bp      mRNA      EST      07-APR-1999
DEFINITION aa22a01.r1 NCI-CGAP-CCBI Homo sapiens cDNA clone IMAGE:813960 5'
            similar to TR:G559703 G559703 MRNA ;, mRNA sequence.
ACCESSION  AA455674
VERSION     AA455674.1 GI:2178450
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 91)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            On Nov 29, 1993 this sequence version replaced gi:503228.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Insert Length: 1221 Std Error: 0.00
            Seq primer: -28m13 rev2 ET from Amersham

```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:28:23 ; Search time 4425.31 Seconds  
(without alignments)  
-308.763 Million cell updates/sec

Title: US-09-215-435-116

Perfect score: 450

Sequence: 1 ctgctcagcgctgacgccc.....acggaaacaagtctaggacag 450

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 821193 seqs, -1518192014 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:

- 1: gb\_ba1:\*\*
- 2: gb\_ba2:\*\*
- 3: gb\_om:\*\*
- 4: gb\_ov:\*\*
- 5: gb\_pat:\*\*
- 6: gb\_ph:\*\*
- 7: gb\_pl1:\*\*
- 8: gb\_pl2:\*\*
- 9: gb\_pri:\*\*
- 10: gb\_pr2:\*\*
- 11: gb\_pr3:\*\*
- 12: gb\_ro:\*\*
- 13: gb\_sts:\*\*
- 14: gb\_sy:\*\*
- 15: gb\_un:\*\*
- 16: gb\_v1:\*\*
- 17: em\_fun:\*\*
- 18: em\_hum1:\*\*
- 19: em\_hum2:\*\*
- 20: em\_in:\*\*
- 21: em\_om:\*\*
- 22: em\_or:\*\*
- 23: em\_ov:\*\*
- 24: em\_pat:\*\*
- 25: em\_ph:\*\*
- 26: em\_pi:\*\*
- 27: em\_ro:\*\*
- 28: em\_sy:\*\*
- 29: em\_un:\*\*
- 30: em\_v1:\*\*
- 31: em\_v2:\*\*
- 32: gb\_htg1:\*\*
- 33: gb\_htg2:\*\*
- 34: gb\_in1:\*\*
- 35: gb\_in2:\*\*
- 36: em\_ba1:\*\*
- 37: em\_ba2:\*\*
- 38: em\_hum3:\*\*
- 39: em\_hum4:\*\*
- 40: gb\_pr4:\*\*
- 41: gb\_htg3:\*\*
- 42: gb\_htg4:\*\*
- 43: gb\_htg5:\*\*

- 44: gb\_htg6:\*\*
- 45: gb\_htg7:\*\*
- 46: em\_htg1:\*\*
- 47: em\_htg2:\*\*
- 48: em\_htg3:\*\*
- 49: em\_hum5:\*\*
- 50: gb\_pl3:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	450	100.0	559	40	AF014955	Homo sapi
2	96	21.3	184976	41	AC008474	Homo sapi
3	83	18.4	173222	42	AC010882	Homo sapi
4	50	11.1	381	12	AF161074	Mus muscu
5	35	7.8	184976	41	AC008474	Homo sapi
6	22	4.9	77673	45	AC016990	Homo sapi
7	22	4.9	219779	44	AC009994	Homo sapi
8	21	4.7	1199	4	GA16534	Gallus gall
9	20	4.4	37017	34	CELF34D8	Caenorhabdi
10	20	4.4	41369	10	CH19F24590	Homo sapi
11	20	4.4	76007	11	AC004503	Homo sapi
12	20	4.4	159941	33	AC006878	Caenorhab
13	20	4.4	162083	40	AC005962	Homo sapi
14	20	4.4	179231	41	AC009549	Homo sapi
15	19	4.2	468	13	G23140	human S15 W
16	19	4.2	2143	11	HOMFRCD	Homo sapien
17	19	4.2	7016	1	BACDAP	Arabidops
18	19	4.2	86001	7	AB016872	Arabidops
19	19	4.2	107541	8	ATAC005499	Arabidops
20	19	4.2	109710	11	HSBA46E17	Human DNA
21	19	4.2	110000	32	PFMAL13P2_1	Continuation (2 of
22	19	4.2	139407	44	AC013290	Homo sapi
23	19	4.2	143700	41	AC009333	Homo sapi
24	19	4.2	147371	11	AC003667	Homo sapi
25	19	4.2	156763	45	AC013284	Homo sapi
26	19	4.2	168237	43	AC010812	Homo sapi
27	19	4.2	178266	44	AC013436	Homo sapi
28	19	4.2	208780	1	BSUB0009	Bacillus su
29	19	4.2	213113	42	AC010100	Homo sapi
30	19	4.2	224719	44	AC007956	Homo sapi
31	18	4.0	178	13	G43987	WTAFC-3614-S
32	18	4.0	178	13	G59861	SHGC-130637
33	18	4.0	1154	8	TA087163	Triticum ae
34	18	4.0	1197	1	AMYF450VD	Amycolata a
35	18	4.0	1216	34	AF052054	Pristionc
36	18	4.0	1260	12	RATGP330	Rat Heymann
37	18	4.0	1478	12	S67967	heparin bin
38	18	4.0	1574	7	SCPDG6	S.cerevisia
39	18	4.0	2768	12	RN45KDA	R.norvegicu
40	18	4.0	2902	12	MUSHP44	Mus musculu
41	18	4.0	3972	9	HUMCDG25C	Human (CDC2
42	18	4.0	3983	7	SCYGR087C	S.cerevisia
43	18	4.0	4191	7	SMPHY1G	S.martensii
44	18	4.0	5577	7	SCYGR088W	S.cerevisia
45	18	4.0	6512	7	NTITS2618	N.tomentosi

ALIGNMENTS

RESULT 1

AF014955

LOCUS

DEFINITION

ACCESSION

VERSION

AF014955

Homo sapiens

TFAR19 mRNA, complete cds.

AF014955.1

GI:2772828

PRI

01-FEB-1999

[illegible]





```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="unknown"
/clone="NH0328P13"
/clone_lib="unknown"
71..91
/rpt_family="AT_rich"
473..761
/rpt_family="Alu"
2142..2301
/rpt_family="Alu"
2302..4181
/rpt_family="L1"
4604..4833
/rpt_family="Alu"
5138..5442
/rpt_family="Alu"
5492..5799
/rpt_family="Alu"
5806..5912
/rpt_family="MER2_type"
5916..6217
/rpt_family="Alu"
6511..6804
/rpt_family="Alu"
6878..6906
/rpt_family="Alu"
6915..7198
/rpt_family="Alu"
7483..7645
8941..9252
/rpt_family="Alu"
9309..9543
/rpt_family="Alu"
9587..9882
/rpt_family="Alu"
9929..9958
/rpt_family="AT_rich"
10038..10148
/rpt_family="MIR"
10149..10552
/rpt_family="MaLR"
10553..10689
/rpt_family="MIR"
11081..11117
/rpt_family="AT_rich"
11351..11626
/rpt_family="Alu"
11627..11646
/rpt_family="(CAAAA)n"
11860..12142
/rpt_family="Alu"
12519..12803
/rpt_family="Alu"
12804..12826
/rpt_family="(TAAAA)n"
12827..12862
/rpt_family="MER1_type"
12863..13163
/rpt_family="Alu"
13164..13198
/rpt_family="MER1_type"
13199..13507
/rpt_family="Alu"
13508..13579
/rpt_family="MER1_type"
13580..13885
/rpt_family="Alu"
13886..13927
/rpt_family="MER1_type"
14344..14629
/rpt_family="Alu"
14849..15105
/rpt_family="Alu"
15106..15415
/rpt_family="Alu"
15416..15739
/rpt_family="Alu"
15740..15761
/rpt_family="AT_rich"
16542..16594
/rpt_family="MIR"
17112..17140
/rpt_family="AT_rich"
17244..17264
/rpt_family="AT_rich"
17796..18087
/rpt_family="Alu"
18170..18456
/rpt_family="Alu"
18457..18750
/rpt_family="Alu"
18780..18875
/rpt_family="Retroviral"
19445..19729
/rpt_family="Alu"
19733..20008
/rpt_family="Alu"
20281..20367
/rpt_family="Alu"
20678..20847
/rpt_family="Alu"
21211..21475
/rpt_family="Alu"
21704..21948
/rpt_family="Alu"
21974..22284
/rpt_family="Alu"
22540..22669
/rpt_family="L2"
22722..23009
/rpt_family="Alu"
24268..24315
/rpt_family="AT_rich"
24330..24442
/rpt_family="Alu"
24448..24589
/rpt_family="Alu"
24614..24856
/rpt_family="Alu"
24858..25137
/rpt_family="MER2_type"
25323..25374
/rpt_family="L2"
25376..25428
/rpt_family="MER2_type"
25430..25546
/rpt_family="Alu"
25547..25974
/rpt_family="MER2_type"
25975..26041
/rpt_family="(TATATG)n"
26089..26200
/rpt_family="MIR"
26311..26516
/rpt_family="MER1_type"
26681..26991
/rpt_family="Alu"
27502..27814
/rpt_family="Alu"
28118..28141
/rpt_family="(A)n"
29443..29749
/rpt_family="Alu"
29753..29878
/rpt_family="L1"
```

```
repeat_region 29948..30089
repeat_region /rpt_family="Alu"
30168..30260

Query Match 18.4% Score 83; DB 42; Length 172322;
Best Local Similarity 100.0%; Pred. No. 1.9e-36;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 190 ttaagtaacttagcactgtgaagcctgaaacactaaagcagtagagattaccttata 249
|||||
Db 142677 TTAAGTAACCTTAGCACTGTGAAGCCTGAAACAACTAAAGCAGTAGAGATTACCTTATA 142736
|||||

Qy 250 cagatggcgaagatatggacaact 272
|||||
Db 142737 CAGATGGCAAGATATGGCAACT 142759
|||||

RESULT 4
AF161074 AF161074 381 bp mRNA ROD 27-JUL-1999
LOCUS Mus musculus TF-1 apoptosis related protein 19 (Tfar19) mRNA,
DEFINITION complete cds.
ACCESSION AF161074
VERSION AF161074.1 GI:5596625
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 381)
AUTHORS Song, Q., Han, W., Liu, H. and Ma, D.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-1999) Department of Immunology, Beijing Medical
University, Xueyuan Road, Beijing 100083, P.R.China
FEATURES
source 1..381
/organism="Mus musculus"
/db_xref="taxon:10090"
1..381
/gene="Tfar19"
1..381
/gene="Tfar19"
/feature="Tfar19"
/codon_start=1
/product="TF-1 apoptosis related protein 19"
/protein_id="AAD45607.1"
/db_xref="GI:5596626"
/translation="MADELEALRQRLAEQLAKHDPGDAACQAEAKOREAEMRNSIL
AOVLDSARARLSNLALVKPEKTKAVENYLLQMARYGOLSGKVEQGLIELEKVSQQ
TEKTTYKFNRRKYWDEDDADI"
BASE COUNT 142 a 71 c 106 g 62 t
ORIGIN

Query Match 11.1% Score 50; DB 12; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1e-17;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 agggagcagaatgagaacagtagcttagcccaagttcttgatcagtc 176
|||||
Db 103 AGGGAGCAGAAATGAGAAACAGTATCTTAGCCCAAGTCTCTGATCAGTC 152
|||||

RESULT 5
AC008474 AC008474 184976 bp DNA HTG 03-SEP-1999
LOCUS Homo sapiens chromosome 19 clone CIT-HSPC-379B2, *** SEQUENCING IN
DEFINITION PROGRESS ***, 47 unordered pieces.
ACCESSION AC008474
VERSION AC008474.1 GI:5686592
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
```

```
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
* NONE: This is a 'working draft' sequence. It currently
* consists of 47 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 614: contig of 614 bp in length
* gap of unknown length
* 615 1318: contig of 704 bp in length
* gap of unknown length
* 1319 1973: contig of 655 bp in length
* gap of unknown length
* 1974 2677: contig of 704 bp in length
* gap of unknown length
* 2678 3326: contig of 649 bp in length
* gap of unknown length
* 3327 3986: contig of 660 bp in length
* gap of unknown length
* 3987 4634: contig of 648 bp in length
* gap of unknown length
* 4635 5261: contig of 627 bp in length
* gap of unknown length
* 5262 5860: contig of 599 bp in length
* gap of unknown length
* 5861 6560: contig of 700 bp in length
* gap of unknown length
* 6561 7265: contig of 705 bp in length
* gap of unknown length
* 7266 8101: contig of 836 bp in length
* gap of unknown length
* 8102 8790: contig of 689 bp in length
* gap of unknown length
* 8791 10137: contig of 1347 bp in length
* gap of unknown length
* 10138 12158: contig of 2021 bp in length
* gap of unknown length
* 12159 13032: contig of 874 bp in length
* gap of unknown length
* 13033 14470: contig of 1438 bp in length
* gap of unknown length
* 14471 16282: contig of 1812 bp in length
* gap of unknown length
* 16283 18442: contig of 2160 bp in length
* gap of unknown length
* 18443 21120: contig of 2678 bp in length
* gap of unknown length
* 21121 24036: contig of 2916 bp in length
* gap of unknown length
* 24037 27253: contig of 3217 bp in length
* gap of unknown length
* 27254 30773: contig of 3520 bp in length
* gap of unknown length
* 30774 34989: contig of 4216 bp in length
* gap of unknown length
* 34990 39988: contig of 4999 bp in length
* gap of unknown length
* 39989 44819: contig of 4831 bp in length
* gap of unknown length
* 44820 49290: contig of 4471 bp in length
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 19  
Unpublished  
2 (bases 1 to 184976)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
www.jgi.doe.gov.



\* \* gap of unknown length  
\* 13263 14169: contig of 907 bp in length  
\* gap of unknown length  
\* 14170 15064: contig of 895 bp in length  
\* gap of unknown length  
\* 15065 15961: contig of 897 bp in length  
\* gap of unknown length  
\* 15962 16956: contig of 895 bp in length  
\* gap of unknown length  
\* 16857 17735: contig of 879 bp in length  
\* gap of unknown length  
\* 17736 18630: contig of 895 bp in length  
\* gap of unknown length  
\* 18631 19497: contig of 867 bp in length  
\* gap of unknown length  
\* 19498 20409: contig of 912 bp in length  
\* gap of unknown length  
\* 20410 21284: contig of 875 bp in length  
\* gap of unknown length  
\* 21285 22175: contig of 891 bp in length  
\* gap of unknown length  
\* 22176 23075: contig of 900 bp in length  
\* gap of unknown length  
\* 23076 23961: contig of 886 bp in length  
\* gap of unknown length  
\* 23962 24859: contig of 898 bp in length  
\* gap of unknown length  
\* 24860 25754: contig of 895 bp in length  
\* gap of unknown length  
\* 25755 26651: contig of 897 bp in length  
\* gap of unknown length  
\* 26652 27550: contig of 899 bp in length  
\* gap of unknown length  
\* 27551 28431: contig of 881 bp in length  
\* gap of unknown length  
\* 28432 29349: contig of 918 bp in length  
\* gap of unknown length  
\* 29350 30253: contig of 904 bp in length  
\* gap of unknown length  
\* 30254 31128: contig of 875 bp in length  
\* gap of unknown length  
\* 31129 32030: contig of 902 bp in length  
\* gap of unknown length  
\* 32031 32885: contig of 855 bp in length  
\* gap of unknown length  
\* 32886 33732: contig of 847 bp in length  
\* gap of unknown length  
\* 33733 34627: contig of 895 bp in length  
\* gap of unknown length  
\* 34628 35519: contig of 892 bp in length  
\* gap of unknown length  
\* 35520 36405: contig of 886 bp in length  
\* gap of unknown length  
\* 36406 37309: contig of 904 bp in length  
\* gap of unknown length  
\* 37310 38217: contig of 908 bp in length  
\* gap of unknown length  
\* 38218 39126: contig of 909 bp in length  
\* gap of unknown length  
\* 39127 40008: contig of 882 bp in length  
\* gap of unknown length  
\* 40009 40902: contig of 894 bp in length  
\* gap of unknown length  
\* 40903 41794: contig of 892 bp in length  
\* gap of unknown length  
\* 41795 42651: contig of 857 bp in length  
\* gap of unknown length  
\* 42652 43561: contig of 910 bp in length  
\* gap of unknown length  
\* 43562 44466: contig of 905 bp in length  
\* gap of unknown length  
\* 44467 45392: contig of 926 bp in length  
\* gap of unknown length

\* 45393 46294: contig of 902 bp in length  
\* gap of unknown length  
\* 46295 47185: contig of 891 bp in length  
\* gap of unknown length  
\* 47186 48049: contig of 864 bp in length  
\* gap of unknown length  
\* 48050 48938: contig of 889 bp in length  
\* gap of unknown length  
\* 48939 49851: contig of 913 bp in length  
\* gap of unknown length  
\* 49852 50721: contig of 870 bp in length  
\* gap of unknown length  
\* 50722 51567: contig of 846 bp in length  
\* gap of unknown length  
\* 51568 52470: contig of 903 bp in length  
\* gap of unknown length  
\* 52471 53371: contig of 901 bp in length  
\* gap of unknown length  
\* 53372 54267: contig of 896 bp in length  
\* gap of unknown length  
\* 54268 55157: contig of 890 bp in length  
\* gap of unknown length  
\* 55158 56062: contig of 905 bp in length  
\* gap of unknown length  
\* 56063 56971: contig of 909 bp in length  
\* gap of unknown length  
\* 56972 57861: contig of 890 bp in length  
\* gap of unknown length  
\* 57862 58763: contig of 902 bp in length  
\* gap of unknown length  
\* 58764 59674: contig of 911 bp in length  
\* gap of unknown length  
\* 59675 60555: contig of 881 bp in length  
\* gap of unknown length  
\* 60556 61425: contig of 870 bp in length  
\* gap of unknown length  
\* 61426 62309: contig of 884 bp in length  
\* gap of unknown length  
\* 62310 63212: contig of 903 bp in length  
\* gap of unknown length  
\* 63213 64165: contig of 953 bp in length  
\* gap of unknown length

Query Match 4.9%; Score 22; DB 45; Length 77673;  
Best Local Similarity 100.0%; Pred. No. 0.18;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 tgaattcaacagaagaagt 371  
|||||  
Db 47829 TGAATTCACAGAGAAAGT 47850

RESULT 7

AC009994 AC009994 219779 bp DNA HTG 07-DEC-1999  
LOCUS Homo sapiens chromosome 17 clone RP11-87N3 map 17, \*\*\* SEQUENCING  
DEFINITION IN PROGRESS \*\*\*, 62 unordered pieces.  
AC009994  
AC009994.3 GI:6532129  
VERSION HTG; HTGS\_PHASE1.  
KEYWORDS human.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 219779)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 17, clone RP11-87N3  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 219779)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Allen,N., Anderson,M.,  
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,  
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,

Fri May 12 12:27:21 2000

```

Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,
Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
Hagos, B., Hearford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,
Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,
Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,
Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
Teschke, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
Direct Submission
Submitted (09-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 7, 1999 this sequence version replaced gi:6102670.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center Project name: L2115
Center Clone name: 87_N_3
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1060: contig of 1060 bp in length
* gap of unknown length
* 1061 2087: contig of 1027 bp in length
* gap of unknown length
* 2088 3173: contig of 1086 bp in length
* gap of unknown length
* 3174 4296: contig of 1123 bp in length
* gap of unknown length
* 4297 5831: contig of 1535 bp in length
* gap of unknown length
* 5832 7257: contig of 1426 bp in length
* gap of unknown length
* 7258 8376: contig of 1119 bp in length
* gap of unknown length
* 8377 9463: contig of 1087 bp in length
* gap of unknown length
* 9464 10771: contig of 1308 bp in length
* gap of unknown length
* 10772 12087: contig of 1316 bp in length
* gap of unknown length
* 12088 13289: contig of 1202 bp in length
* gap of unknown length
* 13290 14949: contig of 1660 bp in length
* gap of unknown length
* 14950 16255: contig of 1306 bp in length
* gap of unknown length
* 16256 17723: contig of 1468 bp in length
* gap of unknown length
* 17724 19390: contig of 1667 bp in length
* gap of unknown length
* 19391 20695: contig of 1305 bp in length
* gap of unknown length
* 20696 22031: contig of 1336 bp in length
* gap of unknown length
* 22032 23614: contig of 1583 bp in length
* gap of unknown length
* 23615 25898: contig of 2284 bp in length
* gap of unknown length
*
* 25899 27605: contig of 1707 bp in length
* gap of unknown length
* 27606 28791: contig of 1186 bp in length
* gap of unknown length
* 28792 30786: contig of 1995 bp in length
* gap of unknown length
* 30787 33274: contig of 2488 bp in length
* gap of unknown length
* 33275 34284: contig of 1010 bp in length
* gap of unknown length
* 34285 35514: contig of 1230 bp in length
* gap of unknown length
* 35515 37802: contig of 2288 bp in length
* gap of unknown length
* 37803 39647: contig of 1845 bp in length
* gap of unknown length
* 39648 41367: contig of 1720 bp in length
* gap of unknown length
* 41368 43395: contig of 2028 bp in length
* gap of unknown length
* 43396 45942: contig of 2547 bp in length
* gap of unknown length
* 45943 47846: contig of 1904 bp in length
* gap of unknown length
* 47847 49854: contig of 2008 bp in length
* gap of unknown length
* 49855 51367: contig of 1513 bp in length
* gap of unknown length
* 51368 53516: contig of 2149 bp in length
* gap of unknown length
* 53517 55949: contig of 2433 bp in length
* gap of unknown length
* 55950 58986: contig of 3037 bp in length
* gap of unknown length
* 58987 61371: contig of 2385 bp in length
* gap of unknown length
* 61372 62516: contig of 1145 bp in length
* gap of unknown length
* 62517 65023: contig of 2507 bp in length
* gap of unknown length
* 65024 68093: contig of 3070 bp in length
* gap of unknown length
* 68094 72034: contig of 3941 bp in length
* gap of unknown length
* 72035 74304: contig of 2270 bp in length
* gap of unknown length
* 74305 76961: contig of 2657 bp in length
* gap of unknown length
* 76962 80151: contig of 3190 bp in length
* gap of unknown length
* 80152 83185: contig of 3034 bp in length
* gap of unknown length
* 83186 86577: contig of 3392 bp in length
* gap of unknown length
* 86578 90674: contig of 4097 bp in length
* gap of unknown length
* 90675 94162: contig of 3488 bp in length
* gap of unknown length
* 94163 100715: contig of 6553 bp in length
* gap of unknown length
* 100716 105211: contig of 4496 bp in length
* gap of unknown length
* 105212 110954: contig of 5743 bp in length
* gap of unknown length
* 110955 118100: contig of 7146 bp in length
* gap of unknown length
* 118101 123095: contig of 4995 bp in length
* gap of unknown length
* 123096 128936: contig of 5841 bp in length
* gap of unknown length
* 128937 136506: contig of 7570 bp in length
* gap of unknown length
* 136507 143211: contig of 6705 bp in length

```

TITLE  
JOURNAL  
COMMENT

```

* 143212 151503: contig of 8292 bp in length
* 151504 162145: contig of 10642 bp in length
* 162146 174699: contig of 12554 bp in length
* 174700 184750: contig of 10051 bp in length
* 184751 199880: contig of 15130 bp in length
* 199881 219779: contig of 19899 bp in length.
FEATURES
  source
    1..219779
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="17"
      /map="17"
      /clone="RP11-87N3"
      /clone_lib="RP11-11 Human Male BAC"
      /base_count 67067 a 43204 c 42715 g 66746 t 47 others

```

## FEATURES

source

```

BASE COUNT 67067 a 43204 c 42715 g 66746 t 47 others
ORIGIN

```

```

Query Match 4.9%; Score 22; DB 44; Length 219779;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 324 ccacaaacagaaaaagacaaca 345
|||||
DB 134400 CCACAAACAGAAAGACACA 134421

```

## RESULT 8

```

LOCUS GGAI6534 1199 bp mRNA VRT 21-SEP-1998
DEFINITION Gallus gallus mRNA for apolipoprotein AIV.
ACCESSION Y16534
VERSION Y16534.1 GI:3645996
KEYWORDS apoAIV gene; apolipoprotein AIV.
SOURCE chicken.
ORGANISM Gallus gallus

```

```

REFERENCE 1 (bases 1 to 1199)
AUTHORS Steinmetz,A., Hermann,M., Nimpf,J., Abersold,R., Ducret,A., Weinberg,R.B. and Schneider,W.J.
TITLE Expression and conservation of apolipoprotein AIV in an avian species
JOURNAL J. Biol. Chem. 273 (17), 10543-10549 (1998)
MEDLINE 98221191
REFERENCE 2 (bases 1 to 1199)
AUTHORS Nimpf,J.
TITLE Direct Submission
JOURNAL Submitted (11-FEB-1998) J. Nimpf, University & Bioenter of Vienna, Dept of Molecular Genetics, Dr. Bohrgasse 9/2, A-1030 Vienna, AUSTRIA

```

## FEATURES

source

gene

CDS

```

1..1199
  /organism="Gallus gallus"
  /db_xref="taxon:9031"
37..1137
  /gene="apoAIV"
37..1137
  /gene="apoAIV"
  /codon_start=1
  /product="apolipoprotein AIV"
  /protein_id="CAA76273.1"
  /db_xref="GI:3645997"
  /db_xref="SPTREMBL:O93601"
  /translation="MSPKALLIVLAVAGTRADVSPDQAVTLVRYFTLGSAAKET
VQLOAQETTKOLNTLQSNLOSANSYAEELQRLVPEATLQALVQDSORLKKQIQ
QELAEIQAQAPVADEVHQIGTINRELQAKLSFYADELRSQVDGTEGELRALEPFA

```

```

TELREKLDQNRDSIQASIGCPYAEIRLQQQIDSSVEGLKGQLTFLADELKEQVAQSVGL
RKGLSPYAEVQDGLNRQLQSLTAQMERAAEELRSRLAASSEMAQUSPLAQELQEA
LRGDAEQQRPLAPLAQQLDERLAQTVEAFRQQAAPISETFQQQLVORLEENKQKLES
GTAGVEDHLDLLEKEVRKVAFLQSTTEQAES"
BASE COUNT 239 a 380 c 427 g 153 t
ORIGIN

```

```

Query Match 4.7%; Score 21; DB 4; Length 1199;
Best Local Similarity 100.0%; Pred. No. 0.53;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 63 gctggcgcgagctgcagggccaa 83
|||||
DB 348 GCTGCCGCGAGCTGCAGGCCAA 368

```

## RESULT 9

CEL54D8

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted by:

Genome Sequencing Center

Department of Genetics, Washington University,

St. Louis, MO 63110, USA, and

Sanger Centre, Hinxton Hall

Cambridge CB10 1RQ, England

e-mail: rw@ematode.wustl.edu and jes@sanger.ac.uk

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is C45C9 ; 3' cosmid is F21H11 (overlaps by 3500

bases). Actual sequence of F54D8 extends from 1 to 37017 of this

sequence.

NOTES:

Coding sequences below are predicted from computer analysis, using

the program GeneFinder (P. Green and L. Hillier, ms in

preparation).

Location/Qualifiers  
1. .37017  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/map="III"  
/strain="Bristol N2"  
1896. .4775  
/gene="F54D8.3"  
join(1896. .1951,2036. .2129,2842. .2970,3018. .3214,  
3435. .3480,3556. .3832,3879. .4625,4687. .4775)  
/gene="F54D8.3"  
/note="highly similar to aldehyde dehydrogenase (ALDH-E2);  
coded for by C.elegans cDNA cm596; coded for by C.  
elegans cDNA cm597; coded for by C.elegans cDNA cm5b1;  
coded for by C.elegans cDNA cm5f10; coded for by C.  
elegans cDNA cm01a9"  
/codon\_start=1  
/protein\_id="AAA20615.1"  
/db\_xref="GI:529223"  
/translation="MFYETSCLELTISNLHMLRSALRATVQARNAGVPPGLSNM  
KPYTGIFINNEVPKASKTETTERINPANGKVLQAVAGDKTDYNAVKAQNAFRIG  
SEWRDASQGVLLNRDLMMEDRIVILASLESNDGKPYAVAINADLPUSIKLRK  
LTAQKNFLNLILFLRRMADNHGKTIPIEDGYTYRHEPVGCGOIIIPNPLLM  
QAKLKPALANGVTVMVAEOTPLSALHVAALTEAGFPDGVVNIIPGYHTAGQAI  
SSHMVDKVAFTSTEVGRVLVMAKAAESNVKVTLELGGKSPNIFADADLNDVSHQA  
NHGLFNQGCCAGSTVEGKIVDDFVARSKELAERKAVIGDPDLATGTPGVODGK  
QVETILKIYAAGKAGQALVTVGAKHGDDGHVFKTIFANVKDQMTIAOEIIFGPMVT  
IIFDPMEELEVERKANNITVGLAAGVMTKIDIKALHIANATRAGSVWVNCYDVFDAAP  
FGFGKSGTGREGLEYGLAYTEVTVTKVPKNS"  
6005. .7127  
/gene="F54D8.2"  
join(6005. .6140,6208. .6359,7029. .7127)  
/gene="F54D8.2"  
/note="similar to cytochrome C oxidase"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAA20614.1"  
/db\_xref="GI:529222"  
/translation="MNRQAQPAVRVYVTKTQKSGSFYGSNNVEGFKESYVPLKQA  
HNASETWKKLFFIASIPCLALTWYAAFKDHHKMSHERPEHYEYAFNVRNKPFPWS  
GNLSLFNKAQEQVPGVGFADREKH"  
complement(7476. .9559)  
/gene="F54D8.4"  
complement(join(7476. .7619,7667. .7888,7940. .8120,  
8474. .8714,8821. .8894,9324. .9559))  
/gene="F54D8.4"  
/note="similar to carbonic anhydrase I"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAA20616.1"  
/db\_xref="GI:529224"  
/translation="MRPECSHFPLFIILTCISPLKSSQYQWSDVDFGGPHFWG  
LVEKDWCKKGLQSDIPQRLILDASVQPSVAENFFLIYITLSKLMRIIF  
TCAHLFACNKKRGNIRIFGAKIAKRAWVYIIIFAGVPLQVYSEVNTQKQVRI  
RIGYSKKPSNVNITSGLYGVYVORIDPHMGKNSEHTINGRRPEVQVLVAY  
NDIPLNFTSASKSPHGLAISLVDPFGTNPENELIKLFIATIASISKDQVQLADFE  
PWRLLPTFRDITTEGSLTPGCKTIVTIIINQPIFIKHEFEWHSYLISNEGAEK  
VPVAPNFRKIQETNNRLVRIINQHWIFLS"  
complement(15320. .15874)  
/gene="F54D8.5"  
complement(15320. .15874)  
/note="similar to ZK1236.3"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAA20617.1"  
/db\_xref="GI:529225"  
/translation="MKSFTMHQIFSNVCSKNTLRHLLTYKTFVRPILYKGVTSVSP  
KKQDKIAYESVQNSFTKLYSRQKVIIRDPPDKSAIORNELENLSTLEARRKAID  
KYLKSLMVLNRYVDITDQFFDKNSKTRKTFEIKWCKTKLRRHFFNRLVLSIKI  
SAMNLLAPHVLCMFNTFTYSSP"  
26759. .32712  
/gene="F54D8.1"

join(26759. .26891,27261. .27796,27846. .27991,28430. .28530,  
28582. .28673,28726. .28876,28822. .28875,28925. .28965,  
29020. .29153,29372. .29660,29792. .30124,30211. .30324,  
30600. .30710,30766. .30864,30932. .31008,31599. .32016,  
32069. .32712)  
/gene="F54D8.1"  
/note="similar to triple helical region of collagens;  
coded for by C.elegans cDNA CEESG54F; coded for by C.  
elegans cDNA CEESA57F"  
/codon\_start=1  
/protein\_id="AAA20613.1"  
/db\_xref="GI:529221"  
/translation="MIAFHRSTAEELSPVSNVEYCDYCSKNTVPGLDAINFCLKQKME  
ENTPVEQODENENTLQIVDSLVIVNSVTNPSENQDQVEMETVSDSEPLEPVP  
PVANFEAPVEEVEGEAEVLVEEVEEVEPAEAEIKSHQRESFMEMFRIPIRDI  
LPDLVFMHSDVDLDEKEVEALVEAKIQDEKAKKAEKSKKEKEVEKPPVSLDEA  
VKTVTGIDALLQIKPNQOIAPIVIPPVPMPPPIVIGMPNAPVATEVLNLVSL  
PLNAKPPPIPIASSYAPNATAPPTIALTDPNVASIGHGKAGSTPHNPNEIPIELHS  
KPDMIAPIDELRNILGAVPVVTIPTPNEELPEQRLILGLPITFVRPMKSLFSL  
GARSCKDRIRKQKDKKVEKLEKSEDSLESEKREPRGGHGRAGRAGRSF  
EATGNALPQGLNKNPLAVKRIQALAKETGMTLEAMEEIEEQNAFEGEDTPEA  
IFHDKLMEKIKRGVDRPGGRRDRSREDPGGVPVDRPAGRSRGIDRPPPGGD  
SVDRPPGGGIGIPGDRPPGGRGVDRPPGSGMPGIDRPPGDRMGDKQRGARGG  
VGERKKRRDFQFGGGGGGGRANVYVEELVYLNKLDVIDMITIDRKEQVDM  
DKIMTHKMNVTMTIIDKRIIMEVAKMENIISNTYKDTINKDMGRMEHIMDM  
KAMRHRTIKMAVNTITNTLLQIKMSFAGTACTGAVSMLLCLVSLPVIYQ  
VHGLDELITTEAMWRLSDQIYMDQFGRVRRQAGYGGYGGYGGSGSPGSPGN  
HGFGPGPGHFPNGTNSNTPLTGVIGVPSVTPHGGPVPINPDGSPSAGPGDKN  
CWTNSCPAGPAGPKGTGHGDPGDPGVPDGEDADDAKAQOQVDDGCTCPAGQ  
GPQSGQFGARGMGARGQARGDGRGSGMGLGPIGPPGAAGEGPTGEPGADV  
EHQIGLPGAGTGPAGSGGEGGEGCGDGRGTAGIAGPPGERGQGEKDDGPNAGASP  
GEEGPGQAQTCPCPQRNTNAVSGNGYRN"

BASE COUNT 12257 a 6360 c 6575 g 11825 t  
ORIGIN

Query Match 4.4%; Score 20; DB 34; Length 37017;  
Best Local Similarity 100.0%; Pred.No.2.4; 0; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0

Oy 321 aagccacaacagaaaaa 340  
|||||  
Db 27015 AAGCCACACACAGAAAAGA 27034

RESULT 10  
CH19F24590/c  
LOCUS CH19F24590 41369 bp DNA PRI 01-APR-1997  
DEFINITION Homo sapiens DNA from chromosome 19-cosmid f24590 containing CAPNS  
and POL2R1, genomic sequence.  
ACCESSION AD001527  
VERSION AD001527.1 GI:1905899  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 41369)  
AUTHORS Garcia,E., Kyle,A., Ramirez,M., Stillwegen,S., Garnes,J.,  
Danganan,L., Bruce,R., Quan,G., Montgomery,M., Ow,D., Kobayashi,A.,  
Olsen,A.O. and Carrano,A.V.  
Sequence analysis of a 1 Mb region in human 19q13.1  
TITLE Sequence analysis of a 1 Mb region in human 19q13.1  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 41369)  
AUTHORS Lamerdin,J.E.  
Direct Submission  
TITLE Submitted (07-FEB-1997) J.E. Lamerdin, Human Genome Center,  
JOURNAL Lawrence Livermore National Laboratory, 7000 East Ave, Livermore,  
CA, USA, 94551 jane@agtl.llnl.gov ow@etornak.llnl.gov  
GSDB:S:1225252.  
COMMENT Gene predictions were accomplished using Xgenail 1.3c  
coupled with local blast comparisons to Genbank, non-redundant



```

protein libraries, and dbEST. Repeat analysis was accomplished
using cross_match against REPBAS.
map-19q13.1 between D19S208 and CAPNS
Human Genome Center
Biology and Biotechnology Research Program
Lawrence Livermore National Laboratory
7000 East Avenue
Livermore, CA 94550 USA.
FEATURES
    source
        1. 41369
            Location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /sub_clone="19"
                /cell_line="UV5HL9-5B"
                /cell_type="fibroblast"
                /clone="f12321"
                /clone_lib="LL19NC02 F2 chromosome 19-specific cosmid
                library"
                /note="constructed at LML from flow-sorted chromosomes
                from hybrid UV5HL9-5B, which carries chromosome 19 as its
                only human chromosome; is adjacent to cosmid R33799 to the
                left (right end of current tiling path)"
            repeat_region
                1. 60
                    /note="repeat match = HSA000392; putative"
                    /rpt_family="Alu"
            repeat_region
                1. 76
                    /note="repeat match = HSA006147; putative"
                    /rpt_family="Alu"
            repeat_region
                183..482
                    /note="repeat match = HSA001424; putative"
                    /rpt_family="Alu"
                    join(1099..1427,2841..2958,3447..3655)
                    /note="comment for location 3447-3655: BLASTX
                    gi|103290|pir||S16356 ovo protein - fruit fly (Drosophila
                    melanogaster), Pval= 3.8e-47"
                    /codon_start=3
                    /protein_id="AAB51180.1"
                    /db_xref="GI:1905900"
                    /translation="ASPPFPFGIVILGRSDSRWKQPLCVESLPGMRAFLVRSRRP
                    OPPNMGHLDPDLRGDAIPGGPLTPVGGKGORRSVTWLFSSDCSSILGGPPAQOSSG
                    VRDPWTARMLTRHLKCHSPVRRHLRCRCGKGFHDALDKRMHTGTGIRPRCSACGG
                    AFTORCSLEHLAKVHGQPASVAYRERREKLHVCDGCGFTSSRPDTYAQHALHRAA"
                    1123..1290
                        /note="predicted exon, grail2exons_human_1.3; frame=2,
                        forward strand, quality=good; putative"
                        1363..1427
                            /note="predicted exon, grail2exons_human_1.3; frame=1,
                            forward strand, quality=excellent; putative"
                            1934..2526
                                /note="repeat match = HSA001529; putative"
                                /rpt_family="Alu"
                                complement(2090..3748)
                                    /gene="POLR21"
                                    /note="comment for location 2090-3748: 3'-UTR from
                                    (Z23102) H.sapiens gene for RNA polymerase II 14.5 kDa
                                    subunit; 3'UTR; 3'-UTR from (Z23102) H.sapiens gene for
                                    RNA polymerase II 14.5 kDa subunit; putative"
                                    complement(2090..5239)
                                        /gene="POLR21"
                                        2841..2958
                                            /gene="POLR21"
                                            /note="predicted exon, grail2exons_human_1.3; frame=2,
                                            forward strand, quality=excellent; putative"
                                            3447..3655
                                                /note="predicted exon, grail2exons_human_1.3; frame=2,
                                                forward strand, quality=excellent; putative"
                                                3447..3655
                                                    /gene="POLR21"
                                                    /note="predicted exon, grail2exons_human_1.3; frame=1,
                                                    forward strand, quality=excellent; putative"
                                                    complement(join(3749..3811,4009..4060,4171..4245,
                                                    4333..4406,4653..4707,4792..4850))
                                                        /gene="POLR21"
                                                        /function="DNA-DEPENDENT RNA POLYMERASE"
                                                        /note="comment for location 4792-4850: exon 1 from

```

```

(Z23102) H.sapiens gene for RNA polymerase II 14.5 kDa
subunit; comment for location 4653-4707: exon 2 from
(Z23102) H.sapiens gene for RNA polymerase II 14.5 kDa
subunit; comment for location 4333-4406: exon 3 from
(Z23102) H.sapiens gene for RNA polymerase II 14.5 kDa
subunit; user-supplied translation (frame
MEPDGTTEPGVGIRFCQECNMMLYPREDKRENRLLYACRNCDFQQAADNSCIYVVKITHEVDEL
THEYDELTIQIIA
DVSQDPTLPTREDHPCQKGHEAVFFQSHSARAEDAMRLYYVCTAPCHGHWTEK;
comment for location 4171-4245: exon 4 from (Z23102)
H.sapiens gene for RNA polymerase II 14.5 kDa subunit;
comment for location 4009-4060: exon 5 from (Z23102)
H.sapiens gene for RNA polymerase II 14.5 kDa subunit;
comment for location 3749-3811: exon 6 from (Z23102)
H.sapiens gene for RNA polymerase II 14.5 kDa subunit;
putative"
/codon_start=1
/product="HUMAN DNA-DIRECTED RNA POLYMERASE II 14.5 KD
SUBUNIT"
/protein_id="AAB51181.1"
/db_xref="GI:1905901"
/translation="MEPDGTTEPGVGIRFCQECNMMLYPREDKRENRLLYACRNCDF
QQAADNSCIYVVKITHEVDELTIQIIADVSDPTLPTREDHPCQKGHEAVFFQSHS
ARAEDAMRLYYVCTAPCHGHWTE"
complement(3752..3814)
/gene="POLR21"
/note="similarity: sp|P36954|RPB9_HUMAN; DNA-DIRECTED RNA
POLYMERASE II 14.2 KD POLYPEPTIDE; putative"
complement(4009..4062)
/gene="POLR21"
/note="similarity: sp|P36954|RPB9_HUMAN; DNA-DIRECTED RNA
POLYMERASE II 14.2 KD POLYPEPTIDE; putative"
complement(4170..4244)
/gene="POLR21"
/note="similarity: sp|P36954|RPB9_HUMAN; DNA-DIRECTED RNA
POLYMERASE II 14.2 KD POLYPEPTIDE; putative"
complement(4332..4406)
/gene="POLR21"
/note="similarity: sp|P36954|RPB9_HUMAN; DNA-DIRECTED RNA
POLYMERASE II 14.2 KD POLYPEPTIDE; putative"
complement(4653..4706)
/gene="POLR21"
/note="similarity: sp|P36954|RPB9_HUMAN; DNA-DIRECTED RNA
POLYMERASE II 14.2 KD POLYPEPTIDE; putative"
complement(4794..4850)
/gene="POLR21"
/note="similarity: sp|P36954|RPB9_HUMAN; DNA-DIRECTED RNA
POLYMERASE II 14.2 KD POLYPEPTIDE; putative"
complement(4851..5239)
/gene="POLR21"
/note="similarity: sp|P36954|RPB9_HUMAN; DNA-DIRECTED RNA
POLYMERASE II 14.2 KD POLYPEPTIDE; putative"
complement(4851..5239)
/note="5'UTR; 5'-UTR from (Z23102) H.sapiens gene for RNA
polymerase II 14.5 kDa subunit; comment for location
4851-5239: 5'-UTR from (Z23102) H.sapiens gene for RNA
polymerase II 14.5 kDa subunit; putative"
5266..5506
/note="predicted exon, grail2exons_human_1.3; frame=2,
forward strand, quality=good; putative"
5555..5668
/note="predicted exon, grail2exons_human_1.3; frame=1,
forward strand, quality=good; putative"
6061..6178
/note="predicted exon, grail2exons_human_1.3; frame=1,
forward strand, quality=excellent; putative"
join(6074..6178,10704..10824,11521..11712,15445..15517,
15662..15776)
/function="unknown"
/note="comment for location 11521-11712: exon 3 similar
to: gi|1469096|gnl|PID|e25598 (Z77663) F53F4.3
[Caenorhabditis elegans], Pval= 9.6e-491; comment for
location 15445-15517: exon 4 similar to:
gi|1469096|gnl|PID|e25598 (Z77663) F53F4.3
[Caenorhabditis elegans], Pval= 9.6e-49; comment for
location 15662-15776: exon 5 similar to:

```



```
/note="GRAIL 2 excellent exon, frame 2"
/complement(46368..46472)
/rpt_family="MIR"
48908..>75715
/gene="BIGH3"
/note="transforming growth factor-beta induced gene
product (BIGH3)"
Join(48955..49088,53662..53760,63957..64021,66234..66394,
66750..66914,67173..67319,69338..69479,72806..73018,
73842..73979,74615..74760,75579..>75715)
/gene="BIGH3"
/note="transforming growth factor-beta induced gene
product (BIGH3)"
/codon_start=1
/evidence-experimental
/product="BIGH3"
/protein_id="AAC08449.1"
/db_xref="GI:2996636"
/translation="MALEFVRLALALALALGPRATLAGPAKSPYQLVLRHLEGRGH
GNVCVQKRVGNTNRKFTCKQWIKRGKSTVISECCPGTEKVPGEKGPAAALP
LSNYETLVGVSTTQLYDTRTEKLEPMGSGSTIFAPSENAWASLPAEVLDSLY
SNVIELNALRYHVMGRVLTDLKHPMTLSMYQNSNIQHHPNGIVTVNCARLL
KADHATNGVHLIDKVIISTITNNIOIIEIDTFELRAAASGLTMEGNGVY
LLAPTNAEFKIPSETINRILGDEALRDLNHLKSAEALVAGLSVETLEGT
LEVCGSDMTINGKALISNKDILATNGVHYIDELLPDSAKTLFELAESVSTAI
DLFQAGLHLSGSERLTLPLNSVFKDGTTPIDAHTRNLRLNHLIKDOLASKYLY
HGQLETLGGKKLRFVYRNSLCIENSCIAAHDKRGYGTLTMDRVLTPPMGTVMYD
LKGNRP"
repeat_region 55224..55552
/rpt_family="Alu"
repeat_region 58141..58334
/rpt_family="MLT1"
repeat_region 58303..58632
/rpt_family="Alu"
repeat_region 58629..58886
/rpt_family="MLT1"
repeat_region 60208..60533
/rpt_family="MER1B"
repeat_region complement(62232..62307)
/rpt_family="MIR"
repeat_region 63480..63559
/rpt_family="MIR"
misc_feature complement(63619..63756)
/note="GRAIL 2 excellent exon, frame 0"
repeat_region complement(64678..64950)
/rpt_family="MER2"
repeat_region 65024..65474
/rpt_family="Alu"
repeat_region 65447..65474
/note="(AAAT)7"
/rpt_type=tandem
repeat_region /rpt_unit=AAAT
69938..70262
/rpt_family="Alu"
repeat_region 71056..71368
/rpt_family="Alu"
misc_feature 71449..71590
/gene="BIGH3"
/note="GRAIL 2 excellent exon, frame 0"
71652..71691
/gene="BIGH3"
/note="GRAIL 2 excellent exon, frame 0"
75793..76002
/note="100% identity B43847 BAC end sequence"
BASE COUNT 20153 a 16885 c 17588 g 21391 t
ORIGIN

/note="GRAIL 2 excellent exon, frame 2"
/complement(46368..46472)
/rpt_family="MIR"
48908..>75715
/gene="BIGH3"
/note="transforming growth factor-beta induced gene
product (BIGH3)"
Join(48955..49088,53662..53760,63957..64021,66234..66394,
66750..66914,67173..67319,69338..69479,72806..73018,
73842..73979,74615..74760,75579..>75715)
/gene="BIGH3"
/note="transforming growth factor-beta induced gene
product (BIGH3)"
/codon_start=1
/evidence-experimental
/product="BIGH3"
/protein_id="AAC08449.1"
/db_xref="GI:2996636"
/translation="MALEFVRLALALALALGPRATLAGPAKSPYQLVLRHLEGRGH
GNVCVQKRVGNTNRKFTCKQWIKRGKSTVISECCPGTEKVPGEKGPAAALP
LSNYETLVGVSTTQLYDTRTEKLEPMGSGSTIFAPSENAWASLPAEVLDSLY
SNVIELNALRYHVMGRVLTDLKHPMTLSMYQNSNIQHHPNGIVTVNCARLL
KADHATNGVHLIDKVIISTITNNIOIIEIDTFELRAAASGLTMEGNGVY
LLAPTNAEFKIPSETINRILGDEALRDLNHLKSAEALVAGLSVETLEGT
LEVCGSDMTINGKALISNKDILATNGVHYIDELLPDSAKTLFELAESVSTAI
DLFQAGLHLSGSERLTLPLNSVFKDGTTPIDAHTRNLRLNHLIKDOLASKYLY
HGQLETLGGKKLRFVYRNSLCIENSCIAAHDKRGYGTLTMDRVLTPPMGTVMYD
LKGNRP"
repeat_region 55224..55552
/rpt_family="Alu"
repeat_region 58141..58334
/rpt_family="MLT1"
repeat_region 58303..58632
/rpt_family="Alu"
repeat_region 58629..58886
/rpt_family="MLT1"
repeat_region 60208..60533
/rpt_family="MER1B"
repeat_region complement(62232..62307)
/rpt_family="MIR"
repeat_region 63480..63559
/rpt_family="MIR"
misc_feature complement(63619..63756)
/note="GRAIL 2 excellent exon, frame 0"
repeat_region complement(64678..64950)
/rpt_family="MER2"
repeat_region 65024..65474
/rpt_family="Alu"
repeat_region 65447..65474
/note="(AAAT)7"
/rpt_type=tandem
repeat_region /rpt_unit=AAAT
69938..70262
/rpt_family="Alu"
repeat_region 71056..71368
/rpt_family="Alu"
misc_feature 71449..71590
/gene="BIGH3"
/note="GRAIL 2 excellent exon, frame 0"
71652..71691
/gene="BIGH3"
/note="GRAIL 2 excellent exon, frame 0"
75793..76002
/note="100% identity B43847 BAC end sequence"
BASE COUNT 20153 a 16885 c 17588 g 21391 t
ORIGIN

Query Match 4.4%; Score 20; DB 11; Length 76007;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 178 gcccgccaggttaagtaa 197

Db 48874 GCCCGGCCAGGTTAAGTAA 48855
|||||
AC006878 159941 bp DNA HTG 26-FEB-1999
Caenorhabditis elegans clone Y54H5, *** SEQUENCING IN PROGRESS ***
8 unordered pieces.
AC006878
AC006878.2 GI:4309916
HTG; HTGS PHASE1
Caenorhabditis elegans.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditidae; Rhabditidae; Pelodidae; Caenorhabditis.
1 (bases 1 to 159941)
Waterston,R.H.
The sequence of Caenorhabditis elegans clone
Unpublished
2 (bases 1 to 159941)
Waterston,R.H.
Direct Submission
Submitted (24-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 1, 1999 this sequence version replaced gi:4263470.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 19303: contig of 19303 bp in length
* 19304 19312: gap of unknown length
* 19313 77334: contig of 58013 bp in length
* 77336 77334: gap of unknown length
* 77335 89842: contig of 12508 bp in length
* 89843 96745: gap of unknown length
* 96746 96745: contig of 6895 bp in length
* 96747 96755: gap of unknown length
* 96756 129022: contig of 32267 bp in length
* 129023 129030: gap of unknown length
* 129031 132373: contig of 3343 bp in length
* 132374 132381: gap of unknown length
* 132382 136331: contig of 3950 bp in length
* 136332 136339: gap of unknown length
* 136340 159941: contig of 23602 bp in length.
FEATURES
source
1..159941
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y54H5"
BASE COUNT 50893 a 28726 c 28118 g 51265 t 939 others
ORIGIN

Query Match 4.4%; Score 20; DB 33; Length 159941;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 321 aagccaacaacagagaaga 340
|||||
Db 62209 AGCCACACACACACACAGA 62190

RESULT 13
AC005962
LOCUS AC005962 162083 bp DNA PRI 24-NOV-1998
DEFINITION Homo sapiens chromosome 17, clone hRPK.506.H_21, complete sequence.
ACCESSION AC005962
```

VERSION	AC005962.1	GI:3924656	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 162083)		
JOURNAL	Birren,B., Linton,L., Nusbaum,C. and Lander,E.		
REFERENCE	Homo sapiens chromosome 17, clone hRPK.506_H_21		
AUTHORS	2 (bases 1 to 162083)		
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,		
	Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,		
	Brown,A., Castle,A., Cerny,J., Collangelo,M., Collins,S.,		
	Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,		
	Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Heaford,A.,		
	Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Kann,L.,		
	Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,		
	Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,		
	McGurk,A., Mckernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,		
	Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,		
	O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,		
	Severy,P., Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,		
	Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,		
	Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,		
	Ye,W.J., Zhao,J. and Zody,M.		
	Direct Submission		
TITLE	Submitted (11-NOV-1998) Whitehead Institute/MIT Center for Genome		
JOURNAL	Research, 320 Charles Street, Cambridge, MA 02141, USA		
REFERENCE	3 (bases 1 to 162083)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,		
	Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,		
	Brown,A., Castle,A., Cerny,J., Collangelo,M., Collins,S.,		
	Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,		
	Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,		
	Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Heaford,A.,		
	Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,		
	Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,		
	McGurk,A., Mckernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,		
	Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,		
	O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,		
	Severy,P., Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,		
	Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,		
	Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,		
	Ye,W.J., Zhao,J. and Zody,M.		
	Direct Submission		
TITLE	Submitted (24-NOV-1998) Whitehead Institute/MIT Center for Genome		
JOURNAL	Research, 320 Charles Street, Cambridge, MA 02141, USA		
COMMENT	On Nov 24, 1998 this sequence version replaced gi:3907449.		
	All repeats were identified using RepeatMasker: Smit, A.F.A. &		
	Green, P. (1996-1997)		
	http://ftp.genome.washington.edu/RM/RepeatMasker.html		
	Only the first 162083 base pairs of this clone are being submitted.		
	The remainder overlaps accession number AC004687 (WICGR project		
	L319).		
FEATURES	Location/Qualifiers		
source	1. 162083		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="hRPK.506_H_21"		
	/clone_lib="RPC1-11 human BAC library"		
	/map="17"		
	/chromosome="17"		
repeat_region	complement(109..159)		
repeat_region	/rpt_family="MIR"		
repeat_region	complement(276..570)		
repeat_region	/rpt_family="AluSg"		
repeat_region	complement(1137..1872)		
repeat_region	/rpt_family="L2"		
repeat_region	complement(1880..2081)		
repeat_region	/rpt_family="MIR"		
repeat_region	complement(2134..2309)		
repeat_region	/rpt_family="L1MC4"		
repeat_region	complement(14183..14381)		
repeat_region	/rpt_family="LTR33"		
repeat_region	14484..14775		
repeat_region	/rpt_family="L2"		
repeat_region	15750..15950		
repeat_region	/rpt_family="MIR"		
repeat_region	complement(16067..16163)		
repeat_region	/rpt_family="MIR"		
repeat_region	complement(16265..16719)		
repeat_region	/rpt_family="L1MB6"		
repeat_region	complement(16769..16981)		
repeat_region	/rpt_family="MER58A"		
repeat_region	17025..17066		
repeat_region	/rpt_family="L2"		
repeat_region	complement(17414..17844)		
repeat_region	/rpt_family="MLT2E"		
repeat_region	complement(17845..19356)		
repeat_region	/rpt_family="HERVL"		
repeat_region	complement(19375..19919)		
repeat_region	/rpt_family="L1MC4"		
repeat_region	19920..20207		
repeat_region	/rpt_family="AluJb"		
repeat_region	complement(20208..20219)		
repeat_region	/rpt_family="L1MC4"		
repeat_region	complement(20220..20512)		
repeat_region	/rpt_family="L1MA5"		
repeat_region	complement(20513..20979)		
repeat_region	/rpt_family="L1MC4"		
repeat_region	complement(21003..21296)		
repeat_region	/rpt_family="AluJo"		
repeat_region	21515..21556		
repeat_region	/rpt_family="MLTIE"		
repeat_region	complement(21647..21935)		
repeat_region	/rpt_family="AluJb"		
repeat_region	complement(22429..22348)		
repeat_region	/rpt_family="purine-rich"		
repeat_region	complement(22349..23197)		
repeat_region	/rpt_family="L1MC4"		
repeat_region	complement(23228..23617)		
repeat_region	/rpt_family="MLT1A1"		
repeat_region	complement(23649..23680)		
repeat_region	/rpt_family="L1MC4"		
repeat_region	complement(23681..23799)		
repeat_region	/rpt_family="FLAM_C"		
repeat_region	complement(23804..23887)		
repeat_region	/rpt_family="AluY"		
repeat_region	complement(23909..24950)		
repeat_region	/rpt_family="L1MC4"		

```
repeat_region complement(24951..25273)
repeat_region /rpt_family="AluSg"
repeat_region complement(25274..25687)
repeat_region /rpt_family="L1MC4"
repeat_region complement(25688..25994)
repeat_region /rpt_family="AluSg"
repeat_region complement(25995..26451)
repeat_region /rpt_family="L1MC4"
repeat_region complement(26457..26538)
repeat_region /rpt_family="MER47"
repeat_region 26529..26606
repeat_region /rpt_family="MER47"
repeat_region complement(26607..26684)
repeat_region /rpt_family="MER47"
repeat_region complement(26688..26813)
repeat_region /rpt_family="L1MEC"
repeat_region complement(26825..26906)
repeat_region /rpt_family="Tigger3(Golem)"
repeat_region 26954..27590
repeat_region /rpt_family="Tigger3b"
repeat_region 27591..27893
repeat_region /rpt_family="AluJb"
repeat_region 27894..28354
repeat_region /rpt_family="Tigger3b"
repeat_region complement(28366..28476)
repeat_region /rpt_family="L1MEC"
repeat_region complement(28477..28777)
repeat_region /rpt_family="AluSx"
repeat_region complement(28778..29579)
repeat_region /rpt_family="L1MEC"
repeat_region 29645..29791
repeat_region /rpt_family="L1MB7"
repeat_region 29888..30130
repeat_region /rpt_family="MLTIF"
repeat_region complement(30151..30274)
repeat_region /rpt_family="MER53"
repeat_region 30280..30386
repeat_region /rpt_family="MLTIF"
repeat_region complement(30387..30700)
repeat_region /rpt_family="AluSx"
repeat_region complement(30703..31000)
repeat_region /rpt_family="AluJo"
repeat_region 31001..31166
repeat_region /rpt_family="MLTIF"
repeat_region complement(31267..31319)
repeat_region /rpt_family="(TAGA)n"
repeat_region 31369..31540
repeat_region /rpt_family="L1MB5"
repeat_region 31546..31632
repeat_region /rpt_family="MLTIF"
repeat_region 31661..31720
```

Query Match 4.4% Score 20; DB 40; Length 162083;

Best Local Similarity 100.0%; Pred. No. 2.7; Mismatches 0; Indels 0; Gaps 0;

QY 350 tgaattcaacagaagaaa 369

|||||

DB 44020 TGAATTCACAGAGAAA 44039

RESULT 14

AC009549

LOCUS

AC009549

DEFINITION

Human sapiens chromosome 11 clone 359\_E10 map 11, \*\*\* SEQUENCING IN

PROGRESS \*\*\* 6 unordered pieces.

AC009549

AC009549.2

HTG: HTGS\_PHASE1.

HTG: HTGS\_PHASE1.

HTG: HTGS\_PHASE1.

HTG: HTGS\_PHASE1.

HTG: HTGS\_PHASE1.

HTG: HTGS\_PHASE1.

HTG: HTGS\_PHASE1.

#### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

1 (bases 1 to 179231)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,  
Castle,A., Cerny,J., Collins,S., Collins,S., Collamore,A.,  
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,  
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,  
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,  
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,  
Karatas,A., Lehoczy,J., Liew,C., Locke,K., McDonald,P.,  
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,  
Meldrim,J., Mollia,M., Morris,W., Morrow,J., Mychaleckyj,J.,  
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,  
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,  
Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,  
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

##### TITLE

##### JOURNAL

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT

##### COMMENT



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 18:51:15 ; Search time 250.34 Seconds  
(without alignments)  
449.734 Million cell updates/sec

Title: US-09-215-435-116  
Perfect score: 450  
Sequence: 1 ctgtctcagcgtgacgccc.....acggagacaagtctaggacag 450

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 311585 seqs, 125096042 residues

Word size : 0

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	298	66.2	347	1 X41341	Human secreted pro
2	19	4.2	423	1 V88755	Est clone HK669. N
3	18	4.0	3836	1 Q53997	Vitamin D hydroxyl
4	18	4.0	7812	1 X13962	Enterococcus faeca
5	17	3.8	397	1 T44938	Partial sequence o
6	17	3.8	399	1 T44929	Partial sequence o
7	17	3.8	399	1 T44933	Partial sequence o
8	17	3.8	399	1 T44939	Partial sequence o
9	17	3.8	537	1 X38629	Breast cancer asso
10	17	3.8	1816	1 Q52676	Partial sequence o
11	17	3.8	2995	1 V70355	Human h-NUMB encod
12	17	3.8	3791	1 T63575	Chicken beta-actin
13	17	3.8	5511	1 T41853	CDNA encoding plas
14	17	3.8	110000	1 X20248_02	Continuation (3 of
15	16	3.6	153	1 T19316	Human gene signat
16	16	3.6	307	1 X40259	Human secreted pro
17	16	3.6	425	1 V87762	EST clone EQ219. N
18	16	3.6	425	1 V86266	EST clone AJ20. Ne
19	16	3.6	430	1 V41440	Nucleotide sequenc
20	16	3.6	507	1 V86393	EST clone AM1083.
21	16	3.6	528	1 X30863	Streptococcus pneu
22	16	3.6	594	1 T98649	DNA encoding a S.
23	16	3.6	647	1 X21020	Polynucleotide seq
24	16	3.6	772	1 V58363	Coding sequence fo
25	16	3.6	819	1 Q22014	Ospa gene variant
26	16	3.6	819	1 Q90712	B. burgdorferi str
27	16	3.6	990	1 X33777	S. aureus coding s
28	16	3.6	1234	1 X14532	H. pylori GHPO 137
29	16	3.6	1235	1 V20666	Human ninjurin 1 e
30	16	3.6	1236	1 T43469	ATM gene exons 32-
31	16	3.6	1514	1 X52239	Protein PRO231 cDN
32	16	3.6	1530	1 N40242	Bradykinin protein
33	16	3.6	1561	1 N40314	Bradykinin protein

34 16 3.6 1745 1 T94656 Carnation flavonoi  
35 16 3.6 1869 1 X33970 Human TIE ligand N  
36 16 3.6 1957 1 T93809 An artificial modi  
37 16 3.6 2219 1 T36646 Acarbose biosynthe  
38 16 3.6 2730 1 T32583 Alfalfa isoflavone  
39 16 3.6 3231 1 T86486 Eya type I cDNA en  
40 16 3.6 3369 1 N81690 Atrial natriuretic  
41 16 3.6 3892 1 Q79354 Sequence of the ex  
42 16 3.6 4107 1 V68109 cDNA encoding the  
43 16 3.6 5922 1 V52400 Streptococcus pneu  
44 16 3.6 5975 1 N91778 Rabbit seletal mus  
45 16 3.6 5975 1 Q87978 Rabbit skeletal ca

## ALIGNMENTS

RESULT 1  
X41341  
ID X41341 standard; cDNA: 347 BP.  
AC X41341;  
DT 17-JUN-1999 (first entry)  
DE Human secreted protein 5; EST SEQ ID NO:285.  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
OS Homo sapiens.  
PN W0906548-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; IB1222.  
PR 01-AUG-1997; US-905135.  
PA (GEST ) GENSET.  
PI Duclert A. Dumas Milne Edwards J, Lacroix B;  
DR WPI: 99-153778/13.  
DR P-PSDB; Y12508.  
PT New nucleic acids encoding human secreted proteins - obtained from  
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
PT kidney, lung, umbilical cord, placenta and colon tissue  
PS Claim 1; Page 634; 824pp; English.  
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins, and encode the proteins given in Y1251 to Y12514,  
CC respectively. The proteins given represent the signal peptide and an  
CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
CC can be used for producing secreted human gene products. They can also  
CC be used to develop products for diagnosis and therapy. The proteins  
CC obtained may have cytokine activity, cell proliferation/differentiation  
CC activity, haematopoiesis regulating activity, tissue growth regulating  
CC activity, reproductive hormone regulating activity, chemotactic/  
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
CC or other activities. The products can be used in forensic, gene therapy  
CC and chromosome mapping procedures. The sequences can also be used for  
CC obtaining corresponding promoter sequences. The nucleic acids encoding  
CC the signal peptide can be used for directing extracellular secretion of  
CC a polypeptide or the insertion of a polypeptide into a membrane, or  
CC importing a polypeptide into a cell.  
SQ Sequence 347 BP; 124 A; 69 C; 98 G; 53 T;

Query Match 66.2%; Score 298; DB 1; Length 347;  
Best Local Similarity 100.0%; Pred. No. 1.9e-134;  
Matches 298; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 agcgtgacgcgcagccatgaggacgagcgttgagcgtgagagacagagctgg 67  
Dd 1 AGGCGTGTAGCGCGGAGCATGTGGGAGGAGCTTGTAGGCGCTGTAGGAGACAGAGCTGG 60  
QY 68 ccgagctgcagcccaaacacgaggatcctggtgatcgcccaacaggaagcaagcaca 127  
Dd 61 CCAGGCTGACGGCCCAACACGGGGATCCTGGTGTATCGGGCCCAACAGGAAGCAAGCACA 120

```

QY 128 gggaagcagaatgagaaacagtagctattagcccaagttctggatcagtcggccgggcca 187
DB 121 GGGAGCAGGAATGAGAAACAGTATCTTATGCCCAAGTTCGGATCAGTCGGCCGGGCCA 180
QY 188 ggttaagtaactagcacttgtaagcctgaaagcctgaaacaaactaaagcagtagagaattacotta 247
DB 181 GGTTAAGTAACTAGCAGTCTGTTAAAGCCTGAAAAAAGCTAAAGCAGTAGAGAAATTACCTTA 240
QY 248 tacagatggcaagataggaacataagtagagaaggtatcagaacaaggtttaataga 305
DB 241 TACAGATGGCAAGATAGGACAACTAAGTAGAGAGGTATCAGAACAAAGGTTTAAATAGA 298

RESULT 2
V88755/c
ID V88755 standard; cDNA; 423 BP.
AC V88755; 1999 (first entry)
DE EST clone HK669.
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS Homo sapiens.
PN W09845437-A2.
PD 15-OCT-1998.
PR 10-APR-1998; U06956.
PR 10-APR-1997; US-837312.
PA (GEMV ) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
PS Claim 1; Page 501; 641pp; English.
CC The present sequence represents an expressed sequence tag (EST), and is
CC a polynucleotide of the invention. The polynucleotides of the invention
CC are all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
SQ Sequence 423 BP; 84 A; 88 C; 101 G; 150 T;

Query Match 4.2%; Score 19; DB 1; Length 423;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 aaaaactaaagcagtaga 236
DB 215 AAAAAGCTAAAGCAGTAGA 197

RESULT 3
Q53997/c
ID Q53997 standard; DNA; 3836 BP.
AC Q53997;
DE 22-JUN-1994 (first entry)
KW Vitamin D hydroxylase gene.
OS Substitution; transformation; 1-apha, 25-dihydroxyvitamin D; ss.
FH Key Location/Qualifiers
FT cds 2300..3496

```

```

FT W09400576-A.
PN 06-JAN-1994.
PF 24-JUN-1993; J00856.
PR 25-JUN-1992; JP-167644.
PI (TAIS ) TAISHO PHARM CO LTD.
PI Adachi T, Beppu T, Hanada K, Horinouchi S, Kawauchi H;
PI Sasaki J;
DR WPI; 94-026212/03.
DR P-PSDB; R47521.
DR DNA encoding vitamin-D hydroxylase - obtd. from actinomycetes,
PT used to produce vitamin-D3 and related cpds.
PS Claim 1; Fig 6-7; 53pp; Japanese.
CC The sequence is that of a gene from actinomycetes which encodes an
CC enzyme which substitutes the H at position 25 of vitamin D cpds.
CC with a hydroxy group (vitamin D hydroxylase). Microorganisms
CC transformed with the gene may be used to manufacture vitamin D
CC derivatives.
SQ Sequence 3836 BP; 571 A; 1401 C; 1313 G; 550 T;

Query Match 4.0%; Score 18; DB 1; Length 3836;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 gtcggccggggccaggtt 191
DB 3367 GTCGCGCCGGGCCAGGTT 3350

RESULT 4
X12962
ID X12962 standard; DNA; 7812 BP.
AC X12962;
DE Enterococcus faecalis genome contig SEQ ID NO:25.
DE Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
KW Enterococcus faecalis.
OS Enterococcus faecalis.
PN W09850555-A2.
PD 12-NOV-1998.
PR 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI; 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1; Page 374-378; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 7812 BP; 2480 A; 1335 C; 1552 G; 2433 T;

Query Match 4.0%; Score 18; DB 1; Length 7812;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



QY 356 tcaacagaagaagtaa 373  
 |||||||  
 Db 3093 TCAACAGAAGAAAGTAA 3110

# RESULT 5

T44938/C  
 ID T44938 standard; DNA; 397 BP.  
 AC T44938;  
 DT 03-JUN-1997 (first entry)  
 DE Partial sequence of gag gene from HIV-1 gp. O strain BCF13.  
 KW Human immunodeficiency virus type 1; HIV-1; envelop; group M; group O;  
 KW C2V3-env; gp41; gag; retrovirus; strain; gp120; hypervariable loop;  
 KW primer; hybridisation; amplification; PCR; polymerase chain reaction;  
 KW immunogen; antibody; ss.  
 OS Human immunodeficiency virus type 1.  
 PN WO9627013-Al.  
 PD 06-SEP-1996.  
 PF 26-FEB-1996; F00294.  
 PR 27-FEB-1995; FR-002236.  
 PA (ASSI-) ASSISTANCE PUBLIQUE HOPITAUX PARIS.  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PI Chaix-Baudier ML, Lousseret-Ajaka I, Ly T, Saragosti S, Simon F;  
 PT New strains of HIV-1 group O, related DNA fragments, peptide(s) and  
 PT antibodies - useful for diagnosis, screening and typing, or as  
 PT immunogens  
 PS Claim 2; Page 43; 71pp; French.  
 CC Human immunodeficiency virus type 1 (HIV-1) strains are currently divided  
 CC into 2 major groups based on the nucleotide sequences of the envelop gene  
 CC (env): group M containing sub-groups A-G, and group O containing the  
 CC strains ANT70 and Wp5180. The invention relates to the discovery of  
 CC partial sequences of the C2V3-env, gp41 and gag genes (see T44907-39 and  
 CC W07329-64). The novel strains have been deposited as retroviruses CNCM  
 CC I-1544 (BCF02 (ESS)), 1543 (BCF01 (FAN)), 1546 (BCF07 (MAN)), 1547 (BCF08  
 CC (NKO)) and 1545 (BCF03 (POC)). The sequence presented here is from the  
 CC strain BCF02 (ESS) and corresponds to a partial sequence of the gag gene.  
 CC The nucleic acids can be used to detect gp O HIV-1 strains by  
 CC hybridisation or (as primers) by gene amplification, also for screening  
 CC and typing of such strains. Peptides encoded by the nucleic acids can  
 CC be used as immunogens to raise Ab for detecting gp. O HIV-1.  
 SQ Sequence 397 BP; 135 A; 86 C; 93 G; 83 T;

Query Match 3.8%; Score 17; DB 1; Length 397;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 ttgacactgttaagcc 215  
 |||||||  
 Db 169 TTAGCAGCTGTGAAGCC 153

# RESULT 6

T44929/C  
 ID T44929 standard; DNA; 399 BP.  
 AC T44929;  
 DT 03-JUN-1997 (first entry)  
 DE Partial sequence of gag gene from HIV-1 gp. O strain BCF02 (ESS).  
 KW Human immunodeficiency virus type 1; HIV-1; envelop; group M; group O;  
 KW C2V3-env; gp41; gag; retrovirus; strain; gp120; hypervariable loop;  
 KW primer; hybridisation; amplification; PCR; polymerase chain reaction;  
 KW immunogen; antibody; ss.  
 OS Human immunodeficiency virus type 1.  
 PN WO9627013-Al.  
 PD 06-SEP-1996.  
 PF 26-FEB-1996; F00294.  
 PR 27-FEB-1995; FR-002236.  
 PA (ASSI-) ASSISTANCE PUBLIQUE HOPITAUX PARIS.  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PI Chaix-Baudier ML, Lousseret-Ajaka I, Ly T, Saragosti S, Simon F;

DR WPI; 96-412779/41.  
 DR P-PSDB; W07354.  
 PT New strains of HIV-1 group O, related DNA fragments, peptide(s) and  
 PT antibodies - useful for diagnosis, screening and typing, or as  
 PT immunogens  
 PS Claim 2; Page 24; 71pp; French.  
 CC Human immunodeficiency virus type 1 (HIV-1) strains are currently divided  
 CC into 2 major groups based on the nucleotide sequences of the envelop gene  
 CC (env): group M containing sub-groups A-G, and group O containing the  
 CC strains ANT70 and Wp5180. The invention relates to the discovery of  
 CC several new strains of HIV-1 which can be placed in group O, based on the  
 CC partial sequences of the C2V3-env, gp41 and gag genes (see T44907-39 and  
 CC W07329-64). The novel strains have been deposited as retroviruses CNCM  
 CC I-1544 (BCF02 (ESS)), 1543 (BCF01 (FAN)), 1546 (BCF07 (MAN)), 1547 (BCF08  
 CC (NKO)) and 1545 (BCF03 (POC)). The sequence presented here is from the  
 CC strain BCF02 (ESS) and corresponds to a partial sequence of the gag gene.  
 CC The nucleic acids can be used to detect gp O HIV-1 strains by  
 CC hybridisation or (as primers) by gene amplification, also for screening  
 CC and typing of such strains. Peptides encoded by the nucleic acids can  
 CC be used as immunogens to raise Ab for detecting gp. O HIV-1.  
 SQ Sequence 399 BP; 131 A; 83 C; 95 G; 90 T;

Query Match 3.8%; Score 17; DB 1; Length 399;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 ttgacactgttaagcc 215  
 |||||||  
 Db 169 TTAGCAGCTGTGAAGCC 153

# RESULT 7

T44933/C  
 ID T44933 standard; DNA; 399 BP.  
 AC T44933;  
 DT 03-JUN-1997 (first entry)  
 DE Partial sequence of gag gene from HIV-1 gp. O strain BCF07 (MAN).  
 KW Human immunodeficiency virus type 1; HIV-1; envelop; group M; group O;  
 KW C2V3-env; gp41; gag; retrovirus; strain; gp120; hypervariable loop;  
 KW primer; hybridisation; amplification; PCR; polymerase chain reaction;  
 KW immunogen; antibody; ss.  
 OS Human immunodeficiency virus type 1.  
 PN WO9627013-Al.  
 PD 06-SEP-1996.  
 PF 26-FEB-1996; F00294.  
 PR 27-FEB-1995; FR-002236.  
 PA (ASSI-) ASSISTANCE PUBLIQUE HOPITAUX PARIS.  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PI Chaix-Baudier ML, Lousseret-Ajaka I, Ly T, Saragosti S, Simon F;  
 PT New strains of HIV-1 group O, related DNA fragments, peptide(s) and  
 PT antibodies - useful for diagnosis, screening and typing, or as  
 PT immunogens  
 PS Claim 2; Page 26; 71pp; French.  
 CC Human immunodeficiency virus type 1 (HIV-1) strains are currently divided  
 CC into 2 major groups based on the nucleotide sequences of the envelop gene  
 CC (env): group M containing sub-groups A-G, and group O containing the  
 CC strains ANT70 and Wp5180. The invention relates to the discovery of  
 CC several new strains of HIV-1 which can be placed in group O, based on the  
 CC partial sequences of the C2V3-env, gp41 and gag genes (see T44907-39 and  
 CC W07329-64). The novel strains have been deposited as retroviruses CNCM  
 CC I-1544 (BCF02 (ESS)), 1543 (BCF01 (FAN)), 1546 (BCF07 (MAN)), 1547 (BCF08  
 CC (NKO)) and 1545 (BCF03 (POC)). The sequence presented here is from the  
 CC strain BCF07 (MAN) and corresponds to a partial sequence of the gag gene.  
 CC The nucleic acids can be used to detect gp O HIV-1 strains by  
 CC hybridisation or (as primers) by gene amplification, also for screening  
 CC and typing of such strains. Peptides encoded by the nucleic acids can  
 CC be used as immunogens to raise Ab for detecting gp. O HIV-1.  
 SQ Sequence 399 BP; 138 A; 85 C; 94 G; 82 T;

```

Query Match      3.8%; Score 17; DB 1; Length 399;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 ttagcacttgtaaagcc 215
    |||||
Db 169 TTAGCACTTGTAAGCC 153

RESULT 8
T44939/c
ID T44939 standard; DNA; 399 BP.
AC T44939;
DT 03-JUN-1997 (first entry)
DE Partial sequence of gag gene from HIV-1 gp. O strain BCF14.
KW Human immunodeficiency virus type 1; HIV-1; envelop; group M; group O;
KW C2V3-env; gp41; gag; retrovirus; strain; gp120; hypervariable loop;
KW primer; hybridisation; amplification; PCR; polymerase chain reaction;
KW immunogen; antibody; ss.
OS Human immunodeficiency virus type 1.
PN W09627013-A1.
PD 06-SEP-1996.
PF 26-FEB-1996; F00294.
PR 27-FEB-1995; FR-002236.
PA (ASSI-) ASSISTANCE PUBLIQUE HOPITAUX PARIS.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PI Chaix-Baudier ML, Loussert-Ajaka I, Ly T, Saragostl S, Simon F;
DR WPI; 96-412779/41.
DR P-PSDB; W07364.
PT New strains of HIV-1 group O, related DNA fragments, peptide(s) and
PT antibodies - useful for diagnosis, screening and typing, or as
PT immunogens
PS Claim 2; Page 43; 71pp; French.
CC Human immunodeficiency virus type 1 (HIV-1) strains are currently divided
CC into 2 major groups based on the nucleotide sequences of the envelop gene
CC (env): group M containing sub-groups A-G, and group O containing the
CC strains ANT70 and MVP180. The invention relates to the discovery of
CC several new strains of HIV-1 which can be placed in group O, based on the
CC partial sequences of the C2V3-env, gp41 and gag genes (see T44907-39 and
CC W07329-64). The novel strains have been deposited as retroviruses CNCM
CC I-1544 (BCF02 (ESS)), 1543 (BCF01 (FAN)), 1546 (BCF07 (MAN)), 1547 (BCF08
CC (NKO)) and 1545 (BCF03 (POC)). The sequence presented here is from the
CC strain BCF14 and corresponds to a partial sequence of the gag gene. The
CC nucleic acids can be used to detect gp O HIV-1 strains by hybridisation
CC or (as primers) by gene amplification, also for screening and typing of
CC such strains. Peptides encoded by the nucleic acids can be used as
CC immunogens to raise Ab for detecting gp. O HIV-1.
SQ Sequence 399 BP; 136 A; 87 C; 96 G; 80 T;

Query Match      3.8%; Score 17; DB 1; Length 399;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 ttagcacttgtaaagcc 215
    |||||
Db 169 TTAGCACTTGTAAGCC 153

RESULT 9
X39629
ID X39629 standard; DNA; 537 BP.
AC X39629;
DT 02-JUL-1999 (first entry)
DE Breast cancer associated gene.
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
OS Homo sapiens.
PN W09904265-A2.
PD 28-JAN-1999.
PF 15-JUL-1998; U14679.
PR 22-JUN-1998; US-102322.

```

```

PR 17-JUL-1997; US-896164.
PR 10-OCT-1997; US-061599.
PR 10-OCT-1997; US-061765.
PR 10-OCT-1997; US-948705.
PR 11-OCT-1997; GB-021697.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, Ohare M, Obata Y, Old LJ,
PI Fireundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI Tureci O;
DR WPI; 99-132448/11.
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
PS Claim 67; Page 384; 787pp; English.
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
SQ Sequence 537 BP; 128 A; 139 C; 160 G; 105 T;

Query Match      3.8%; Score 17; DB 1; Length 537;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 376 gactctgtagaagatga 392
    |||||
Db 339 GACTCTGATGAGATGA 355

RESULT 10
Q52676/c
ID Q52676 standard; DNA; 1816 BP.
AC Q52676;
DT 03-MAY-1994 (first entry)
DE Partial sequence of Xenopus F-spondin.
KW Thrombospondin; TSR; Thrombospondin type I repeat; F-spondin;
KW neurodevelopment; nerve; axon; adhesion; outgrowth; ss.
OS Gallus gallus.
PH Key Location/Qualifiers
FT cds 2..1705
FT /tag= a
FT /product= F-spondin.
PN W09320196-A.
PD 14-OCT-1993.
PF 02-APR-1993; U03164.
PR 02-APR-1992; US-862021.
PA (UYCO ) UNIV COLUMBIA NEW YORK.
PI Jessell TM, Klar A;
DR WPI; 93-336904/42.
DR P-PSDB; R44243.
PT New vertebrate F-spondin protein - used for attaching nerve cells
PT to a matrix, stimulating growth of nerve cells or regenerating
PT nerve cells
PS Claim 1; Page 57-61; 103pp; English.
CC F-spondin is useful for adhesion and outgrowth of axons. It can be
CC used for attaching nerve cells to a matrix, stimulating growth of
CC nerve cells or regenerating nerve cells. F-spondin nucleic acid
CC can be used to develop probes to study neurodevelopment. The
CC antibodies to the F-spondin can be used for determining the
CC localisation of the protein in the nervous system and in assessing
CC its function.
SQ Sequence 1816 BP; 533 A; 392 C; 499 G; 392 T;

```

transgenic pig; transgenic mouse; antibody mediated rejection;	PF	26-APR-1995; U05798.
hyperacute rejection; antigen reducing enzyme;	PR	27-APR-1995; US-430908.
alpha(1,2)fucosyltransferase; beta-actin; promoter; ss.	PA	(AFY-) AFFYMAX TECHNOLOGIES NV.
Gallus sp.	PI	Baruch DI, Howard RJ, Pasloske BL;

DR WPI: 96-497376/49.  
DR P-PSDB: W00385.  
PT New Plasmodium falciparum erythrocyte membrane proteins - used to  
PT develop products for the diagnosis, treatment or prevention of  
PT malaria parasite infections  
PS Disclosure: Figure 12: 149pp; English.  
CC A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte  
CC membrane protein 1 (PfEMP1) or active fragments or analogues of that  
CC protein can be used in the treatment or prevention of symptoms of a  
CC malaria parasite infection. The polypeptides can inhibit, block or  
CC reverse the sequestration of erythrocytes in patients suffering from  
CC malaria. Nucleic acids derived from the PfEMP1 gene can be used as  
CC probes and primers to identify a Plasmodium falciparum parasite, the  
CC primers used to generate characteristic amplification patterns from  
CC different P. falciparum strains. Antibodies specifically  
CC immunoreactive with the PfEMP1 polypeptide or its fragments may be  
CC used in diagnosis of malaria infection. This sequence encodes a  
CC truncated version of the PfEMP1 protein of the MC type of Plasmodium  
CC falciparum encoded by a cDNA clone. A full length genomic clone  
CC coding sequence is described in T41852.  
SQ Sequence 5511 BP; 2150 A; 913 C; 1189 G; 1259 T;

Query Match 3.8%; Score 17; DB 1; Length 5511;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 ctgatgaagatgacgat 396  
|||||  
DB 5172 CTGATGAAGATGACGAT 5188

RESULT 14  
X20248\_02/c  
Continuation (3 of 10) of X20248 from base 200001 (Borrelia burgdorferi polynucleotide s  
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
WP Fragment Name Begin End  
WP X20248\_00 1 110000  
WP X20248\_01 100001 210000  
WP X20248\_02 200001 310000  
WP X20248\_03 300001 410000  
WP X20248\_04 400001 510000  
WP X20248\_05 500001 610000  
WP X20248\_06 600001 710000  
WP X20248\_07 700001 810000  
WP X20248\_08 800001 910000  
WP X20248\_09 900001 910715

Query Match 3.8%; Score 17; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 tcttagcccaagtctg 168  
|||||  
DB 37973 TCTTAGCCCAAGTCTG 37957

RESULT 15  
T19316/c  
ID T19316 standard; cDNA to mRNA; 153 BP.  
AC T19316;  
DT 28-JUN-1996 (first entry)  
DE Human gene signature HUMGS00338.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN W09514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.

PA (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI: 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1; Page 357; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 153 BP; 33 A; 24 C; 36 G; 59 T;

Query Match 3.6%; Score 16; DB 1; Length 153;  
Best Local Similarity 100.0%; Pred. No. 94;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 tgaaaaaactaaagca 231  
|||||  
DB 83 TGAAAAAACTAAAGCA 68

Search completed: May 1, 2000, 18:51:40  
Job time: 18784 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:04:50 ; Search time 181.17 Seconds  
(without alignments)  
297.375 Million cell updates/sec

Title: US-09-215-435-116  
Perfect score: 450  
Sequence: 1 ctgctcagcgtgaocg.....acggaacaagtctagacag 450

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 214294 seqs, 59861574 residues

Word size : 0

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PTUS9\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	3.8	1341	5	US-09-032-372-9
C 2	17	3.8	1816	1	US-07-862-021B-13
C 3	17	3.8	1816	2	US-08-313-288B-13
C 4	17	3.8	1816	6	PCT-US93-03164-13
5	17	3.8	2307	3	US-08-967-101-28
6	17	3.8	2307	4	US-08-592-541-28
7	16	3.6	819	2	US-08-320-161-9
8	16	3.6	885	2	US-08-628-291-1
9	16	3.6	885	4	US-09-128-722-1
C 10	16	3.6	1190	6	PCT-US91-02626-8
11	16	3.6	1280	4	US-08-628-291-15
12	16	3.6	1280	4	US-09-128-722-15
13	16	3.6	1957	1	US-08-295-060-3
14	16	3.6	2219	2	US-08-606-322-1
15	16	3.6	2730	2	US-08-339-129-1
C 16	16	3.6	3231	1	US-08-195-152-1
17	16	3.6	4107	3	US-08-813-940-24
18	16	3.6	5962	7	5386025-5
19	16	3.6	5975	1	US-08-404-354B-1
20	16	3.6	5975	1	US-08-314-083B-1
21	16	3.6	5975	2	US-08-435-675B-1
22	16	3.6	5975	2	US-08-336-257A-3
C 23	15	3.3	17	3	US-08-332-766A-94
C 24	15	3.3	211	1	US-08-222-177A-22
C 25	15	3.3	421	3	US-08-332-766A-25
C 26	15	3.3	586	2	US-08-385-590A-3

27	15	3.3	696	1	US-08-181-271A-11	Sequence 11, Appl
28	15	3.3	696	1	US-08-449-315-11	Sequence 11, Appl
29	15	3.3	696	1	US-08-444-803-11	Sequence 11, Appl
30	15	3.3	696	1	US-08-449-043-11	Sequence 11, Appl
31	15	3.3	696	2	US-08-456-265A-11	Sequence 11, Appl
32	15	3.3	696	2	US-08-455-416-11	Sequence 11, Appl
33	15	3.3	696	2	US-08-455-244-11	Sequence 11, Appl
34	15	3.3	696	2	US-08-454-876-11	Sequence 11, Appl
35	15	3.3	696	3	US-08-457-364-11	Sequence 11, Appl
36	15	3.3	696	3	US-08-456-262-11	Sequence 11, Appl
37	15	3.3	696	3	US-08-456-240-11	Sequence 11, Appl
38	15	3.3	696	3	US-08-455-736-11	Sequence 11, Appl
39	15	3.3	696	4	US-08-971-217-11	Sequence 11, Appl
40	15	3.3	882	2	US-08-628-291-3	Sequence 3, Appl
41	15	3.3	882	4	US-09-128-722-3	Sequence 3, Appl
C 42	15	3.3	897	4	US-08-486-663A-19	Sequence 19, Appl
43	15	3.3	1014	4	US-09-066-074-1	Sequence 1, Appl
44	15	3.3	1014	4	US-08-555-912A-1	Sequence 1, Appl
45	15	3.3	1204	2	US-08-628-291-11	Sequence 11, Appl

## ALIGNMENTS

RESULT 1  
US-09-032-372-9  
; Sequence 9, Application US/09032372  
; Patent No. 6008337  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Yue, Henry  
; APPLICANT: Lal, Preeti  
; TITLE OF INVENTION: CELL CYCLE RELATED PROTEINS  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/032.372  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0478 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0355  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1341 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: BRAITUT21  
; CLONE: 2522306





```
; FILING DATE: 19930402
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELE: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1816 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1705
PCT-US93-03164-13
```

```
Query Match 3.8%; Score 17; DB 6; Length 1816;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 65 tggcgcagctgcaggcc 81
      |||||||
DB 650 TGGCCGAGCTGCAGGCC 634
```

```
RESULT 5
US-08-967-101-28
; Sequence 28, Application US/08967101
; Patent No. 5840540
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,101
; FILING DATE: 10-NOV-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2307 base pairs
```

```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-967-101-28
```

```
Query Match 3.8%; Score 17; DB 3; Length 2307;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 109 caacaggaagcaaacga 125
      |||||||
DB 1290 CAACAGGAAGCAAGCA 1306
```

```
RESULT 6
US-08-592-541-28
; Sequence 28, Application US/08592541
; Patent No. 5986054
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2307 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-592-541-28
```

```
Query Match 3.8%; Score 17; DB 4; Length 2307;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 109 caacaggaagcaaacga 125
      |||||||
DB 1290 CAACAGGAAGCAAGCA 1306
```

```
RESULT 7
US-08-320-161-9
; Sequence 9, Application US/08320161
; Patent No. 5747294
; GENERAL INFORMATION:
; APPLICANT: Flavell, Richard A.
```

APPLICANT: Kantor, Fred S.  
APPLICANT: Barthold, Stephen W.  
APPLICANT: Fikrig, Erol  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
PREVENTION AND DIAGNOSIS OF LYME DISEASE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: James F. Haley, Jr., Esq. c/o Fish & Neave  
STREET: 875 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10022-6250  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/320,161  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/682,355  
FILING DATE:  
APPLICATION NUMBER: US 538,969  
FILING DATE: 15-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 602,551  
FILING DATE: 26-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: YU-100 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 715-0600  
TELEFAX: (212) 715-0673  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 819 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..819  
US-08-320-161-9

Query Match 3.6%; Score 16; DB 2; Length 819;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 tcaacagagaagaagt 371  
Db 361 TCAACAGAGAAGAAGT 376

RESULT 8  
US-08-628-291-1  
Sequence 1, Application US/08628291  
Patent No. 5801031  
GENERAL INFORMATION:  
APPLICANT: GALIVAN, JOHN H.  
APPLICANT: RYAN, THOMAS J.  
APPLICANT: YAO, RONG  
APPLICANT: NIMEC, ZENIA  
TITLE OF INVENTION: GAMMA GLUTAMYL HYDROLASE  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP  
COUNTRY: USA

ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/628,291  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: TIMIAN, SUSAN J.  
REGISTRATION NUMBER: 34,103  
REFERENCE/DOCKET NUMBER: 20894/150  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 716-263-1636  
TELEFAX: 716-263-1600  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 885 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-628-291-1

Query Match 3.6%; Score 16; DB 2; Length 885;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 atacagatggcgaagt 262  
Db 587 ATACAGATGGCAAGT 602

RESULT 9  
US-09-128-722-1  
Sequence 1, Application US/09128722  
Patent No. 5962235  
GENERAL INFORMATION:  
APPLICANT: Galivan, John H  
APPLICANT: Ryan, Thomas J  
APPLICANT: Yao, Rong  
APPLICANT: Nimec, Zenia  
TITLE OF INVENTION: Gamma Glutamyl Hydrolase  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jaeckle Fleischmann & Mugel, LLP  
STREET: 39 State Street  
CITY: Rochester  
STATE: New York  
COUNTRY: US  
ZIP: 14614-1310  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,722  
FILING DATE: 04-AUG-1998  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/628,291  
FILING DATE: 05-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Braham, Susan J  
REGISTRATION NUMBER: 34,103  
REFERENCE/DOCKET NUMBER: 87681.98R196  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 716-262-3640  
TELEFAX: 716-262-4133

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 885 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-09-128-722-1

Query Match 3.6%; Score 16; DB 4; Length 885;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 atacagatgcaagat 262  
|||||  
DB 587 ATACAGATGCAAGAT 602

RESULT 10  
PCT-US91-02626-8/c  
; Sequence 8, Application PC/TUS9102626  
; GENERAL INFORMATION:  
; APPLICANT: Zarling, David A.  
; APPLICANT: Senda, Elissa P.  
; APPLICANT: Green, Christopher J.  
; TITLE OF INVENTION: Process for Nucleic Acid Hybridization  
; TITLE OF INVENTION: and Amplification  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Swiss  
; STREET: P.O. Box 60850  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US91/02626  
; FILING DATE: 19910405  
; CLASSIFICATION: 435.6  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/520,321  
; FILING DATE: 07-MAY-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dehlinger, Peter J.  
; REGISTRATION NUMBER: 28,006  
; REFERENCE/DOCKET NUMBER: P-2579/8255-001.41  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 323-8302  
; TELEFAX: (415) 323-8306  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1190 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: AQUASPIRILLUM MAGNETOTACTICUM  
; INDIVIDUAL ISOLATE: PARTIAL SEQUENCE OF THE RecA GENE  
PCT-US91-02626-8

Query Match 3.6%; Score 16; DB 6; Length 1190;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 381 tgatgaatgacgat 396  
|||||  
DB 679 TGATGAATGACGAT 664

RESULT 11  
US-08-628-291-15  
; Sequence 15, Application US/08628291  
; Patent No. 5801031  
; GENERAL INFORMATION:  
; APPLICANT: GALIVAN, JOHN H.  
; APPLICANT: RYAN, THOMAS J.  
; APPLICANT: YAO, RONG  
; APPLICANT: NIMEC, ZENIA  
; TITLE OF INVENTION: GAMMA GLUTAMYL HYDROLASE  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP  
; COUNTRY: USA  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/628,291  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: TIMIAN, SUSAN J.  
; REGISTRATION NUMBER: 34,103  
; REFERENCE/DOCKET NUMBER: 20894/150  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 716-263-1636  
; TELEFAX: 716-263-1600  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1280 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
US-08-628-291-15

Query Match 3.6%; Score 16; DB 2; Length 1280;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 atacagatgcaagat 262  
|||||  
DB 718 ATACAGATGCAAGAT 733

RESULT 12  
US-09-128-722-15  
; Sequence 15, Application US/09128722  
; Patent No. 5962235  
; GENERAL INFORMATION:  
; APPLICANT: Galivan, John H.  
; APPLICANT: Ryan, Thomas J.  
; APPLICANT: YAO, RONG  
; APPLICANT: Nimec, Zenia  
; TITLE OF INVENTION: Gamma Glutamyl Hydrolase  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jaecle Fleischmann & Mugel, LLP  
; STREET: 39 State Street  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: US  
; ZIP: 14614-1310

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,722  
FILING DATE: 04-AUG-1998  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/628,291  
FILING DATE: 05-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Braman, Susan J  
REGISTRATION NUMBER: 34,103  
REFERENCE/DOCKET NUMBER: 87681.98R196  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 716-262-3640  
TELEFAX: 716-262-4133  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1280 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-09-128-722-15

Query Match 3.6%; Score 16; DB 4; Length 1280;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 atacagatggcaagat 262  
Db 718 ATACAGATGGCAAGAT 733

RESULT 13  
US-08-295-060-3  
Sequence 3, Application US/08295060  
Patent No. 5659123  
GENERAL INFORMATION:  
APPLICANT: VAN RIE, Jeroen  
APPLICANT: JANSSENS, Stefan  
APPLICANT: PERFEROEN, Marinix  
TITLE OF INVENTION: NEW DIABROTICA TOXINS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/295,060  
FILING DATE: 26-AUG-1994  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 010830-052  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1957 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 13..1947  
US-08-295-060-3

Query Match 3.6%; Score 16; DB 1; Length 1957;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 ctggccgagctgcagg 79  
Db 403 CTGCCGAGCTGCAGG 418

RESULT 14  
US-08-606-322-1  
Sequence 1, Application US/08606322  
Patent No. 5753501  
GENERAL INFORMATION:  
APPLICANT: Crueger, Anneliese; Piepersberg,  
APPLICANT: Wolfgang; Distler, Jurgen; and  
APPLICANT: Stratmann, Ansgar  
TITLE OF INVENTION: ACARBOSE BIOSYNTHESIS GENES FROM  
TITLE OF INVENTION: ACTINOPLANES sp., PROCESS FOR THE ISOLATION  
TITLE OF INVENTION: THEREOF AND THE USE THEREOF  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SPRUNG HORN KRAMER & WOODS  
STREET: 660 White Plains Road  
CITY: Tarrytown  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10591-5144  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 mb  
MEDIUM TYPE: storage  
COMPUTER: Bravo 3/25s  
OPERATING SYSTEM: DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/606,322  
FILING DATE: 23-FEB-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 195 07 214.6  
FILING DATE: 02-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurt G. Briscoe  
REGISTRATION NUMBER: 33,141  
REFERENCE/DOCKET NUMBER: Bayer 9537-KGB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (914) 332-1700  
TELEFAX: (914) 332-1844  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2219 Nucleotides  
TYPE: Nucleic Acid  
STRANDEDNESS: Double  
TOPOLOGY: Linear  
MOLECULE TYPE: Genomic DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Actinoplanes sp. SE 50/110

US-08-606-322-1

Query Match 3.6%; Score 16; DB 2; Length 2219;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 gatcctgtgtatgcgg 106  
|||||  
DB 41 GATCCTGTGTATGCGG 56

## RESULT 15

US-08-339-129-1

; Sequence 1, Application US/08339129  
; Patent No. 5750399  
; GENERAL INFORMATION:  
; APPLICANT: Dixon, Richard A.  
; APPLICANT: Paiva, Nancy L.  
; APPLICANT: Ommen, Abraham  
; TITLE OF INVENTION: Isoflavone Reductase Promoter  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Richards, Medlock & Andrews  
; STREET: 1201 Elm Street, Suite 4500  
; CITY: Dallas  
; STATE: Texas  
; COUNTRY: US  
; ZIP: 75270-2197  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/339,129  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hansen, Eugenia S.  
; REGISTRATION NUMBER: 31,966  
; REFERENCE/DOCKET NUMBER: NOBF B35969  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 214-939-4500  
; TELEFAX: 214-939-4600  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2730 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-339-129-1

Query Match 3.6%; Score 16; DB 2; Length 2730;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 305 aaatccttaaaaaagt 320  
|||||  
DB 404 AATCCTTAAAAAGT 419

Search completed: May 1, 2000, 16:04:55  
Job time: 18360 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 19:59:49 ; Search time 4088.29 Seconds  
(without alignments)  
-377.456 Million cell updates/sec

Title: US-09-215-435-116  
Perfect score: 450  
Sequence: 1 ctgctcagcgtgaacgcg.....acggaacaagtcttagacag 450

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 5265378 seqs, -1714608768 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 10530736

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Pending\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/pna/US092C\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US081A\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US081B\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US081C\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US081D\_COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US082A\_COMB.seq.\*  
9: /cgn2\_6/ptodata/1/pna/US082B\_COMB.seq.\*  
10: /cgn2\_6/ptodata/1/pna/US082C\_COMB.seq.\*  
11: /cgn2\_6/ptodata/1/pna/US083A\_COMB.seq.\*  
12: /cgn2\_6/ptodata/1/pna/US083B\_COMB.seq.\*  
13: /cgn2\_6/ptodata/1/pna/US084A\_COMB.seq.\*  
14: /cgn2\_6/ptodata/1/pna/US084B\_COMB.seq.\*  
15: /cgn2\_6/ptodata/1/pna/US084C\_COMB.seq.\*  
16: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*  
17: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*  
18: /cgn2\_6/ptodata/1/pna/US087A\_COMB.seq.\*  
19: /cgn2\_6/ptodata/1/pna/US087B\_COMB.seq.\*  
20: /cgn2\_6/ptodata/1/pna/US087C\_COMB.seq.\*  
21: /cgn2\_6/ptodata/1/pna/US088A\_COMB.seq.\*  
22: /cgn2\_6/ptodata/1/pna/US088B\_COMB.seq.\*  
23: /cgn2\_6/ptodata/1/pna/US088C\_COMB.seq.\*  
24: /cgn2\_6/ptodata/1/pna/US088D\_COMB.seq.\*  
25: /cgn2\_6/ptodata/1/pna/US089A\_COMB.seq.\*  
26: /cgn2\_6/ptodata/1/pna/US089B\_COMB.seq.\*  
27: /cgn2\_6/ptodata/1/pna/US089C\_COMB.seq.\*  
28: /cgn2\_6/ptodata/1/pna/US089D\_COMB.seq.\*  
29: /cgn2\_6/ptodata/1/pna/US089E\_COMB.seq.\*  
30: /cgn2\_6/ptodata/1/pna/US089F\_COMB.seq.\*  
31: /cgn2\_6/ptodata/1/pna/US089G\_COMB.seq.\*  
32: /cgn2\_6/ptodata/1/pna/US089H\_COMB.seq.\*  
33: /cgn2\_6/ptodata/1/pna/US089I\_COMB.seq.\*  
34: /cgn2\_6/ptodata/1/pna/US090A\_COMB.seq.\*  
35: /cgn2\_6/ptodata/1/pna/US090B\_COMB.seq.\*  
36: /cgn2\_6/ptodata/1/pna/US090C\_COMB.seq.\*  
37: /cgn2\_6/ptodata/1/pna/US090D\_COMB.seq.\*  
38: /cgn2\_6/ptodata/1/pna/US090E\_COMB.seq.\*  
39: /cgn2\_6/ptodata/1/pna/US091A\_COMB.seq.\*  
40: /cgn2\_6/ptodata/1/pna/US091B\_COMB.seq.\*  
41: /cgn2\_6/ptodata/1/pna/US091C\_COMB.seq.\*  
42: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq.\*  
43: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq.\*

44: /cgn2\_6/ptodata/1/pna/US092C\_COMB.seq.\*  
45: /cgn2\_6/ptodata/1/pna/US092D\_COMB.seq.\*  
46: /cgn2\_6/ptodata/1/pna/US092E\_COMB.seq.\*  
47: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq.\*  
48: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq.\*  
49: /cgn2\_6/ptodata/1/pna/US093C\_COMB.seq.\*  
50: /cgn2\_6/ptodata/1/pna/US093D\_COMB.seq.\*  
51: /cgn2\_6/ptodata/1/pna/US093E\_COMB.seq.\*  
52: /cgn2\_6/ptodata/1/pna/US093F\_COMB.seq.\*  
53: /cgn2\_6/ptodata/1/pna/US094A\_COMB.seq.\*  
54: /cgn2\_6/ptodata/1/pna/US094B\_COMB.seq.\*  
55: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq.\*  
56: /cgn2\_6/ptodata/1/pna/US6001A\_COMB.seq.\*  
57: /cgn2\_6/ptodata/1/pna/US6001B\_COMB.seq.\*  
58: /cgn2\_6/ptodata/1/pna/US6002A\_COMB.seq.\*  
59: /cgn2\_6/ptodata/1/pna/US6002B\_COMB.seq.\*  
60: /cgn2\_6/ptodata/1/pna/US6002C\_COMB.seq.\*  
61: /cgn2\_6/ptodata/1/pna/US6003A\_COMB.seq.\*  
62: /cgn2\_6/ptodata/1/pna/US6003B\_COMB.seq.\*  
63: /cgn2\_6/ptodata/1/pna/US6004A\_COMB.seq.\*  
64: /cgn2\_6/ptodata/1/pna/US6004B\_COMB.seq.\*  
65: /cgn2\_6/ptodata/1/pna/US6004C\_COMB.seq.\*  
66: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq.\*  
67: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq.\*  
68: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq.\*  
69: /cgn2\_6/ptodata/1/pna/US6008A\_COMB.seq.\*  
70: /cgn2\_6/ptodata/1/pna/US6008B\_COMB.seq.\*  
71: /cgn2\_6/ptodata/1/pna/US6008C\_COMB.seq.\*  
72: /cgn2\_6/ptodata/1/pna/US6009A\_COMB.seq.\*  
73: /cgn2\_6/ptodata/1/pna/US6009B\_COMB.seq.\*  
74: /cgn2\_6/ptodata/1/pna/US6010A\_COMB.seq.\*  
75: /cgn2\_6/ptodata/1/pna/US6010B\_COMB.seq.\*  
76: /cgn2\_6/ptodata/1/pna/US6011A\_COMB.seq.\*  
77: /cgn2\_6/ptodata/1/pna/US6011B\_COMB.seq.\*  
78: /cgn2\_6/ptodata/1/pna/US6012A\_COMB.seq.\*  
79: /cgn2\_6/ptodata/1/pna/US6012B\_COMB.seq.\*  
80: /cgn2\_6/ptodata/1/pna/US6013A\_COMB.seq.\*  
81: /cgn2\_6/ptodata/1/pna/US6013B\_COMB.seq.\*  
82: /cgn2\_6/ptodata/1/pna/US6014A\_COMB.seq.\*  
83: /cgn2\_6/ptodata/1/pna/US6014B\_COMB.seq.\*  
84: /cgn2\_6/ptodata/1/pna/US6014C\_COMB.seq.\*  
85: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq.\*  
86: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq.\*  
87: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq.\*  
88: /cgn2\_6/ptodata/2/pna/PCT\_NEW\_COMB.seq.\*  
89: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB.seq.\*  
90: /cgn2\_6/ptodata/2/pna/US07\_NEW\_COMB.seq.\*  
91: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq.\*  
92: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*  
93: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	450	100.0	450	42	US-09-215-435-116	Sequence 116, App
2	450	100.0	548	86	US-60-164-285-6488	Sequence 6488, App
3	450	100.0	559	42	US-09-215-435-516	Sequence 516, App
4	450	100.0	568	45	US-09-289-768-27719	Sequence 27719, A
5	450	100.0	568	48	US-09-332-782-28739	Sequence 28739, A
6	450	100.0	577	24	US-08-870-870-65	Sequence 65, Appl
7	450	100.0	577	24	US-08-870-870A-65	Sequence 65, Appl
8	450	100.0	589	87	US-60-172-373-13544	Sequence 13544, A
9	369	82.0	421	53	US-09-431-517-13611	Sequence 13611, A
10	369	82.0	421	53	US-09-431-517-14582	Sequence 14582, A
11	361	80.2	626	87	US-60-172-373-25470	Sequence 25470, A
12	344	76.4	420	53	US-09-431-517-13610	Sequence 13610, A
13	337	74.9	338	48	US-09-332-782-6058	Sequence 6058, Ap





Db	301	atagaatactcttaaaaaagtaagcccaacaaacagaaaagacaacacagtgaattccaac	360
QY	361	agaagaaaagtaatgactctgatgaatgacgatgcgattattgtaactacaagtgtcacaga	420
Db	361		
Db	361	agaagaaaagtaatggactctgatgaatgacgatgcgattattgtaactacaagtgtcacaga	420
QY	421	ctagaacttaacggaacaagtctaggacag	450
Db	421		
Db	421	ctagaacttaacggaacaagtctaggacag	450
 RESULT 4			
US-09-289-768-27719			
; Sequence 27719, Application US/09289768			
; GENERAL INFORMATION:			
; APPLICANT: Hyseq, Inc.			
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED			
; FILE REFERENCE: 20411-765			
; CURRENT APPLICATION NUMBER: US/09/289,768			
; CURRENT FILING DATE: 1999-04-08			
; NUMBER OF SEQ ID NOS: 39996			
; SOFTWARE: FastSeq for Windows Version 3.0			
; SEQ ID NO 27719			
; LENGTH: 568			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: misc_feature			
; LOCATION: (1)..(568)			
; OTHER INFORMATION: n = A,T,C or G			
US-09-289-768-27719			
 Query Match 100.0%; Score 450; DB 45; Length 568;			
Best Local Similarity 100.0%; Pred. No. 2.3e-232;			
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps			
QY	1	ctgctcagcgctgacgccgagccatggcgacgagagcttgagcgctgaggagacag	60
Db	36	ctgctcagcgctgacgccgagccatggcgacgagagcttgagcgctgaggagacag	95
QY	61	aggctggccgagctgcaggcccaaacacgvgggatcctgtgtgatcgcgcccaacaggaa	120
Db	96	aggctggccgagctgcaggcccaaacacgvgggatcctgtgtgatcgcgcccaacaggaa	155
QY	121	aagcacagggaacagaaaaatgaaaaacagtattcttagcccagttctgtagtcagtcggcc	180
Db	156	aagcacagggaacagaaaaatgaaaaacagtattcttagcccagttctgtagtcagtcggcc	215
QY	181	cgggcaggtttaaagtaacttagcacitctaaagcctgaaaaaactaaagcagtagagaat	240
Db	216	cgggcaggtttaaagtaacttagcacitctaaagcctgaaaaaactaaagcagtagagaat	275
QY	241	tacctatacagatgccaagatatgacaactaatgagaaggtatcagaaacaaggttta	300
Db	276	tacctatacagatgccaagatatgacaactaatgagaaggtatcagaaacaaggttta	335
QY	301	atagaaatcctttaaaaaagtaagcccaacaaacagaaaagacaacacagtgaattccaac	360
Db	336	atagaaatcctttaaaaaagtaagcccaacaaacagaaaagacaacacagtgaattccaac	395
QY	361	agaagaaaactaatggactctgatgaatgacgatgcgattattgtaactacaagtgtcacaga	420
Db	396	agaagaaaactaatggactctgatgaatgacgatgcgattattgtaactacaagtgtcacaga	455
QY	421	ctagaacttaacggaacaagtctaggacag	450
Db	456	ctagaacttaacggaacaagtctaggacag	485
 RESULT 5			
US-09-332-782-28739			

```
; Sequence 28739, Application US/09332782A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/332,782A
; CURRENT FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28739
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(568)
; OTHER INFORMATION: n = A,T,C or G
US-09-332-782-28739

Query Match      100.0%; Score 450; DB 48; Length 568;
Best Local Similarity 100.0%; Pred. No. 2.3e-232;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgtctcagcgctgacccgagccatgagcagagagagcttgagcgctgagagagacg 60
DB 36 ctgtctcagcgctgacccgagccatgagcagagagagcttgagcgctgagagagacg 95
QY 61 aggtggtccagctgacggccaaacacagggatcctgtgtagcgcccaacaggaagca 120
DB 96 aggtggtccagctgacggccaaacacagggatcctgtgtagcgcccaacaggaagca 155
QY 121 aagcacagggagcagcaaaatgagaaacagctatcttagcccaagttctggatcgccg 180
DB 156 aagcacagggagcagcaaaatgagaaacagctatcttagcccaagttctggatcgccg 215
QY 181 cgggcccaggttaagtaacttagcactgttaagcctgaaacactaaacagtagagaat 240
DB 216 cgggcccaggttaagtaacttagcactgttaagcctgaaacactaaacagtagagaat 275
QY 241 taccttatacagatggcaagataggaactaagtgagagaggtatcagaacaagggtta 300
DB 276 taccttatacagatggcaagataggaactaagtgagagaggtatcagaacaagggtta 335
QY 301 atagaaatccttaaaaaagtagcaacaaacacagaaagacaaacagtagtaattcaac 360
DB 336 atagaaatccttaaaaaagtagcaacaaacacagaaagacaaacagtagtaattcaac 395
QY 361 aagaagaaagtaagtgactctgtagaagatgacgattattgaactacaagtgctcacaga 420
DB 396 aagaagaaagtaagtgactctgtagaagatgacgattattgaactacaagtgctcacaga 455
QY 421 ctagaacttaacggaaacagtcaggacag 450
DB 456 ctagaacttaacggaaacagtcaggacag 485

RESULT 6
US-08-870-870-65
; Sequence 65, Application US/08870870
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN REGULATORY PROTEINS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
```

```
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870.870
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0300 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 577 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAITUT02
; CLONE: 2095728
US-08-870-870-65

Query Match      100.0%; Score 450; DB 24; Length 577;
Best Local Similarity 100.0%; Pred. No. 2.3e-232;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgtctcagcgctgacgagccatgagcagagagagcttgagcgctgagagagacg 60
DB 48 CTGCTCCAGCGCTGACGCCGAGCCATGGCGGAGAGCTTGAGCGCTGAGGAGACAG 107
QY 61 aggtggtccagctgacggccaaacacagggatcctgtgtagcgcccaacaggaagca 120
DB 108 AGGCTGGCCGAGCTGCAGGCCAACACAGGGGATCCTTGGTGTGCGGCCCAACAGGAAGCA 167
QY 121 aagcacagggagcagcaaaatgagaaacagtagtatcttagcccaagttctgtagcagtcgccc 180
DB 168 AAGCACAGGGAAGCAGAAAAATGAGAAACAGTATCTTAGCCCAAGTTCTGGATCATGCGGCC 227
QY 181 cgggcccaggttaagtaacttagcactgttaagcctgaaacactaaacagtagagaat 240
DB 228 CGGCGCCAGGTTAAGTAAGTAACTTAGCAGTTGTAAGCCCTGAAAAAACTAAAGCAGTAGAGAAT 287
QY 241 taccttatacagatggcaagataggaactaagtgagagaggtatcagaacaagggtta 300
DB 288 TACCTTATACAGATGGCAAGATATGGACAACATAAGTGAGAGATATCAGAACAAAGGTTTA 347
QY 301 atagaaatccttaaaaaagtagcaacaaacacagaaagacaaacagtagtaattcaac 360
DB 348 ATAGAAATCCTTAAAAAAGTAGCCAAACAAACAGAAAAAGACAAACAGTAGTAATTCAAC 407
QY 361 aagaagaaagtaagtgactctgtagaagatgacgattattgaactacaagtgctcacaga 420
DB 408 AGAAGAAAAGTAAGTGGACTCTGATGAGATGACGATTATTGAACTACAGTGTCTCAGAGA 467
QY 421 ctagaacttaacggaaacagtcaggacag 450
DB 468 CTAGAAGTTAACGGAACAAAGTCTAGGACAG 497

RESULT 7
```

US-08-870-870A-65  
; Sequence 65, Application US/08870870A  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Shah, Purvi  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: NEW HUMAN REGULATORY PROTEINS  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/870,870A  
; FILING DATE: June 6, 1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0300 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 577 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: BRAITUT02  
; CLONE: 2095728  
US-08-870-870A-65

Query Match 100.0%; Score 450; DB 24; Length 577;  
Best Local Similarity 100.0%; Pred. No. 2.3e-232;  
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ctgctccagcgctgacgcccagccatgaggacgagagcttgagcgctgagcgagacag 60  
Db 48 CTGCTCCAGCGCTGACGCCGACCCATGCGCGAGAGAGCTTGAGCGCTGAGGAGACAG 107  
Qy 61 aggtggtggcgagctgagcgcccaaacacggtggtgctggtgagcgcccaacaggaagca 120  
Db 108 AGGCTGGCGGAGCTGCAGGCCAACACAGGGGATCTGTGTGATGCGGCCCAACAGGAAGCA 167  
Qy 121 aagcacagggag 180  
Db 168 AAGCACAGGGAAG 227  
Qy 181 cggtggcaggttaagtaactgacactgtaagcctgaaagcctgaaagcctgaaagcctgaa 240  
Db 228 CGGGCCAGGTTAAGTAACCTTACACTTGTAAAGCCTTGAAGAACTAAGCAGTAGAGAT 287  
Qy 241 taccctatacagatggcagagatgagacactaagtgagaaggtatcagagaacaggttta 300  
Db 288 TACCTTATACAGATGCAAGATATGCAACACTAAGTGAGAAGGTATCAGAACAAAGTTTA 347

Qy 301 atagaaatccttaaaaaaagtaagcccaaacagagaaagacaaacagagtaaatccaac 360  
Db 348 ATAGAAATCCTTAAAAAAGTAAGCCCAACACAGAAAGAGACAAACAGTGAATTCAC 407  
Qy 361 agaagaaaagtaatggactctctgtagagatgacgattattgaactacaagtgcgcacaga 420  
Db 408 AGAAGAAAAGTAATGGACTCTGTATGAAGATGACGATTATTGAACATAAAGTGCTCACAGA 467  
Qy 421 ctagaacttaacggagaaacaaagtctaggacag 450  
Db 468 CTAGAACTTAACGGAGAACAAAGTCTAGGACAG 497  
RESULT 8  
US-60-172-373-13544  
; Sequence 13544, Application US/60172373  
; GENERAL INFORMATION:  
; APPLICANT: Morris, MacDonald  
; APPLICANT: Lal, Preeti  
; APPLICANT: Diep, Dinh  
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using  
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polym  
; FILE REFERENCE: GX-0006 P  
; CURRENT APPLICATION NUMBER: US/60/172,373  
; CURRENT FILING DATE: 1999-12-16  
; NUMBER OF SEQ ID NOS: 25,772  
; SOFTWARE: PERL Program  
; SEQ ID NO 13544  
; LENGTH: 589  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No: 235725.6  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 8  
; OTHER INFORMATION: a, t, c, g, or other  
US-60-172-373-13544

Query Match 100.0%; Score 450; DB 87; Length 589;  
Best Local Similarity 100.0%; Pred. No. 2.3e-232;  
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ctgctccagcgctgacgcccagccatgaggacgagagcttgagcgctgagcgagacag 60  
Db 60 ctgctccagcgctgacgcccagccatgaggacgagagcttgagcgctgagcgagacag 119  
Qy 61 aggtggtggcgagctgacgcccagccatgaggacgagagcttgagcgctgagcgagacag 120  
Db 120 aggtggtggcgagctgacgcccagccatgaggacgagagcttgagcgctgagcgagacag 179  
Qy 121 aagcacagggag 180  
Db 180 aagcacagggag 239  
Qy 181 cggtggcaggttaagtaactgacactgtaagcctgaaagcctgaaagcctgaaagcctgaa 240  
Db 240 cggtggcaggttaagtaactgacactgtaagcctgaaagcctgaaagcctgaaagcctgaa 299  
Qy 241 taccctatacagatggcagagatgagacactaagtgagaaggtatcagagaacaggttta 300  
Db 300 taccctatacagatggcagagatgagacactaagtgagaaggtatcagagaacaggttta 359  
Qy 301 atagaaatccttaaaaaaagtaagcccaaacagagaaagacaaacagagtaaatccaac 360  
Db 360 atagaaatccttaaaaaaagtaagcccaaacagagaaagacaaacagagtaaatccaac 419  
Qy 361 agaagaaaagtaatggactctctgtagaagatgacgattattgaactacaagtgcgcacaga 420  
Db 420 agaagaaaagtaatggactctctgtagaagatgacgattattgaactacaagtgcgcacaga 479

QY 421 ctgaacttaacggagcaaacagtctaggacag 450  
|||||  
Db 480 ctgaacttaacggagcaaacagtctaggacag 509

## RESULT 9

US-09-431-517-13611  
; Sequence 13611, Application US/09431517  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-751CON1  
; CURRENT APPLICATION NUMBER: US/09/431,517  
; CURRENT FILING DATE: 1999-11-01  
; EARLIER APPLICATION NUMBER: US 09/170,294  
; EARLIER FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 31760  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 13611  
; LENGTH: 421  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)-(421)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-431-517-13611

Query Match 82.0%; Score 369; DB 53; Length 421;

Best Local Similarity 100.0%; Pred. No. 1.3e-188; Indels 0; Gaps 0;  
Matches 369; Conservative 0; Mismatches 0;

QY 33 cgaggagcttgagcgctgagagacagagctggcgcgagctgcagggcccaaacacgaggga 92  
|||||  
Db 13 cgaggagcttgagcgctgagagacagagctggcgcgagctgcagggcccaaacacgaggga 72  
|||||  
QY 93 tctgtgtagtcgcccacagagcaaacagagcagggagcaggaatgagaaacagtat 152  
|||||  
Db 73 tctgtgtagtcgcccacagagcaaacagagcagggagcaggaatgagaaacagtat 132  
|||||  
QY 153 cttagcccaagtctgtagtcagtcgcccgcgaggttaagttaacttagcacttgaataa 212  
|||||  
Db 133 cttagcccaagtctgtagtcagtcgcccgcgaggttaagttaacttagcacttgaataa 192  
|||||  
QY 213 gctgaaaaaactaaagcagtagagaattacattacagatggcaagatatggacaact 272  
|||||  
Db 193 gctgaaaaaactaaagcagtagagaattacattacagatggcaagatatggacaact 252  
|||||  
QY 273 aagtgaaggtatcagaacaaggtttaatagaatactttaaagaaagtaagcccaaacac 332  
|||||  
Db 253 aagtgaaggtatcagaacaaggtttaatagaatactttaaagaaagtaagcccaaacac 312  
|||||  
QY 333 agaaaagcaacaacagtagtaattcaacagagaagaaagtaagtgaactctgataagatga 392  
|||||  
Db 313 agaaaagcaacaacagtagtaattcaacagagaagaaagtaagtgaactctgataagatga 372  
|||||  
QY 393 cgattattg 401  
|||||  
Db 373 cgattattg 381

## RESULT 10

US-09-431-517-14582  
; Sequence 14582, Application US/09431517  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-751CON1  
; CURRENT APPLICATION NUMBER: US/09/431,517  
; CURRENT FILING DATE: 1999-11-01

; EARLIER APPLICATION NUMBER: US 09/170,294  
; EARLIER FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 31760  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 14582  
; LENGTH: 421  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)-(421)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-431-517-14582

Query Match 82.0%; Score 369; DB 53; Length 421;

Best Local Similarity 100.0%; Pred. No. 1.3e-188; Indels 0; Gaps 0;  
Matches 369; Conservative 0; Mismatches 0;

QY 33 cgaggagcttgagcgctgagagacagagctggcgcgagctgcagggcccaaacacgaggga 92  
|||||  
Db 13 cgaggagcttgagcgctgagagacagagctggcgcgagctgcagggcccaaacacgaggga 72  
|||||  
QY 93 tctgtgtagtcgcccacagagcaaacagagcagggagcaggaatgagaaacagtat 152  
|||||  
Db 73 tctgtgtagtcgcccacagagcaaacagagcagggagcaggaatgagaaacagtat 132  
|||||  
QY 153 cttagcccaagtctgtagtcagtcgcccgcgaggttaagttaacttagcacttgaataa 212  
|||||  
Db 133 cttagcccaagtctgtagtcagtcgcccgcgaggttaagttaacttagcacttgaataa 192  
|||||  
QY 213 gctgaaaaaactaaagcagtagagaattacattacagatggcaagatatggacaact 272  
|||||  
Db 193 gctgaaaaaactaaagcagtagagaattacattacagatggcaagatatggacaact 252  
|||||  
QY 273 aagtgaaggtatcagaacaaggtttaatagaatactttaaagaaagtaagcccaaacac 332  
|||||  
Db 253 aagtgaaggtatcagaacaaggtttaatagaatactttaaagaaagtaagcccaaacac 312  
|||||  
QY 333 agaaaagcaacaacagtagtaattcaacagagaagaaagtaagtgaactctgataagatga 392  
|||||  
Db 313 agaaaagcaacaacagtagtaattcaacagagaagaaagtaagtgaactctgataagatga 372  
|||||  
QY 393 cgattattg 401  
|||||  
Db 373 cgattattg 381

## RESULT 11

US-60-172-373-25470  
; Sequence 25470, Application US/60172373  
; GENERAL INFORMATION:  
; APPLICANT: Morris, MacDonald  
; APPLICANT: Lal, Preeti  
; APPLICANT: Diep, Dinh  
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using  
; FILE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polym  
; FILE REFERENCE: GX-0006 P  
; CURRENT APPLICATION NUMBER: US/60/172,373  
; CURRENT FILING DATE: 1999-12-16  
; NUMBER OF SEQ ID NOS: 25,772  
; SOFTWARE: PERL Program  
; SEQ ID NO 25470  
; LENGTH: 626  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID No: 235725.4  
US-60-172-373-25470

Query Match 80.2%; Score 361; DB 87; Length 626;



```
; APPLICANT: Drake, Jim
; APPLICANT: Drmanac, Radoje
; APPLICANT: Engleman, Carrie
; APPLICANT: Faulkner, Brandy
; APPLICANT: Fox, Melvin
; APPLICANT: Garcia, Veronica
; APPLICANT: Giedt, Gretchen
; APPLICANT: Jomek, Leni
; APPLICANT: Jones, Lee
; APPLICANT: Kita, David
; APPLICANT: Kofler, Janette
; APPLICANT: Labat, Ivan
; APPLICANT: Lee, Won-Jae
; APPLICANT: Lomelli, Michelle
; APPLICANT: Nguyen, Hong
; APPLICANT: Nguyen, Linh
; APPLICANT: Nguyen, Lynne
; APPLICANT: Nguyen, Phuong
; APPLICANT: Nogra, Margie
; APPLICANT: Ojeda, Jesse
; APPLICANT: Palencia, Servando
; APPLICANT: Raisi, Fariba
; APPLICANT: Randhwa, Gurpreet
; APPLICANT: Sabourieh, Hannah
; APPLICANT: Sidhu, Navjwan
; APPLICANT: Smith, Benjamin
; APPLICANT: Smythe, Ashleigh
; APPLICANT: Tkach, Joe
; APPLICANT: Tulpule, Mukul
; APPLICANT: Verna, Ron
; APPLICANT: Wachter, Adam
; APPLICANT: Wu, James
; APPLICANT: Yim, Kenneth
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 754CIP
; CURRENT APPLICATION NUMBER: US/09/515,694
; CURRENT FILING DATE: 2000-02-29
; EARLIER FILING DATE: 1999-06-14
; EARLIER APPLICATION NUMBER: 09/332,782
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 09/181,430
; NUMBER OF SEQ ID NOS: 21027
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6058
; LENGTH: 338
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-515-694-6058

Query Match      74.9%; Score 337; DB 92; Length 338;
Best Local Similarity 100.0%; Pred. No. 2.5e-171;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 cttgagcgctgagagacagagctggccgagctgcagcccaaacacgggagctcctggt 99
Db 1 cttgagcgctgagagacagagctggccgagctgcagcccaaacacgggagctcctggt 60
QY 100 gatcgcccccaacagggaagcaagcaggggaagcagagaatgagaacacagtattctagcc 159
Db 61 gatcgcccccaacagggaagcaagcaggggaagcagagaatgagaacacagtattctagcc 120
QY 160 caagtcttgatcagtcgccccggccgaggttaagttaacttagcactgttaaacctgaa 219
Db 121 caagtcttgatcagtcgccccggccgaggttaagttaacttagcactgttaaacctgaa 180
QY 220 aaaaactaagcagtagagaattaccttatacagatgcagagatgcacaactaagtga 279
Db 181 aaaaactaagcagtagagaattaccttatacagatgcagagatgcacaactaagtga 240
QY 280 aaggtatcagacaaggtttaataagaatccttaaaaaagtaagcccaacacagaaaaag 339
;

; APPLICANT: Drake, Jim
; APPLICANT: Drmanac, Radoje
; APPLICANT: Engleman, Carrie
; APPLICANT: Faulkner, Brandy
; APPLICANT: Fox, Melvin
; APPLICANT: Garcia, Veronica
; APPLICANT: Giedt, Gretchen
; APPLICANT: Jomek, Leni
; APPLICANT: Jones, Lee
; APPLICANT: Kita, David
; APPLICANT: Kofler, Janette
; APPLICANT: Labat, Ivan
; APPLICANT: Lee, Won-Jae
; APPLICANT: Lomelli, Michelle
; APPLICANT: Nguyen, Hong
; APPLICANT: Nguyen, Linh
; APPLICANT: Nguyen, Lynne
; APPLICANT: Nguyen, Phuong
; APPLICANT: Nogra, Margie
; APPLICANT: Ojeda, Jesse
; APPLICANT: Palencia, Servando
; APPLICANT: Raisi, Fariba
; APPLICANT: Randhwa, Gurpreet
; APPLICANT: Sabourieh, Hannah
; APPLICANT: Sidhu, Navjwan
; APPLICANT: Smith, Benjamin
; APPLICANT: Smythe, Ashleigh
; APPLICANT: Tkach, Joe
; APPLICANT: Tulpule, Mukul
; APPLICANT: Verna, Ron
; APPLICANT: Wachter, Adam
; APPLICANT: Wu, James
; APPLICANT: Yim, Kenneth
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-752CON1
; CURRENT APPLICATION NUMBER: US/09/359,922
; CURRENT FILING DATE: 1999-07-22
; EARLIER APPLICATION NUMBER: US 09/205,155
; EARLIER FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 13203
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4808
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-359-922-4808

Query Match      73.3%; Score 330; DB 50; Length 590;
Best Local Similarity 99.8%; Pred. No. 1.5e-167;
Matches 450; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ctgtccagcgctgacgcccagccatggggacgagagcttgagcgctgagagacag 60
Db 538 CTGCTCCAGCGCTGACGCCGAGCCATGGCGGACGAGAGCTTGAGCGCTGAGGACAG 479
QY 61 aggtcgccgagctgcagcccaacacacggggtcctgggtatcgcccccaacagggaagca 120
Db 478 AGGTGCGCGAGCTGCAGGCCAACACACGGGGATCCTGGTGTATCGGCCCAACAGGAAGCA 419
QY 121 aagcacagggaagcagaaaaagagaacagtatcttagcccaagttctggatcagtcggcc 180
Db 418 AAGCACAGGGAAGCAGANAATGAGAACAGTATCTTAGCCCCAAGTTCTGGATCAGTCGCC 359
QY 181 cgggcccaggttaagttaacttagcactgttaagccttaaacctgaaaaactaaagcagtagagaat 240
Db 358 CGGGCCAGGTTAAGTAACCTTAGCAGCTGTAAGCCCTGAAAAAACTAAAGCAGTAGAGAA 299
QY 241 taccttatacagatggcaagatatgacaa-ctaagtgagaaggtatcagacaaggttt 299
Db 298 TACCTTATACAGATGGCAAGATATGGACACACCTAAGTGAGAGGTATCAGAACAGGTTT 239
QY 300 aatagaaatccttaaaaaagtaagcccaacaaacagacagaaaaacacacagtgaaattcaa 359
Db 238 AATAGAAATCCTTAAAAAAGTAGCCCAACAAACAGAAAAGACAAACACAGTGAATTC 179
QY 360 cagaagaaaaagtaagtgcctctgatgaagatgacgattattgaactacagtgctcacag 419
Db 178 CAGAAGAAAAAGTAAATGACTCTGATGAAGATGACGATTATTGAACATACAGTCTCACAG 119
QY 420 actagaacttaacggaacagctagacag 450
Db 118 ACTAGAACTTAAACGGAACAAGTCTAGGACAG 88

Search completed: May 1, 2000, 19:59:52
Job time: 21976 sec
```



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 14:44:32 ; Search time 3022.95 Seconds  
(without alignments)  
562.050 Million cell updates/sec

Title: us-09-215-435-116  
Perfect score: 450  
Sequence: 1 ctgctccagcgtgacgcg.....acggaacaagtctaggacag 450

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4538634 seqs, 1887831982 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 9077368

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*

44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_gss1:\*  
80: gb\_gss2:\*  
81: gb\_gss3:\*  
82: gb\_gss4:\*  
83: em\_gss1:\*  
84: em\_gss2:\*  
85: em\_gss3:\*  
86: em\_gss4:\*  
87: gb\_gss5:\*  
88: gb\_gss6:\*  
89: gb\_gss7:\*  
90: gb\_gss8:\*  
91: gb\_gss9:\*  
92: em\_gss5:\*  
93: em\_gss6:\*  
94: em\_gss7:\*  
95: em\_gss8:\*  
96: em\_gss9:\*  
97: em\_gss10:\*  
98: em\_gss11:\*  
99: gb\_gss10:\*  
100: gb\_gss11:\*  
101: em\_gss12:\*  
102: gb\_gss12:\*  
103: gb\_gss13:\*  
104: gb\_gss14:\*  
105: gb\_gss15:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
1	450	100.0	476	33	AA452724	AA452724	zx39d11.r



KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE 1 (bases 1 to 573)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1404977.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Ilian Kirsch, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 592 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 446.  
FEATURES  
source  
1. .573  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1144701"  
/clone\_lib="NCI-CGAP\_Col0"  
/tissue\_type="colon tumor RER"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: p7T3D-Pac (Pharmacia) with a  
modified polylinker; 1st strand cDNA was prepared from  
RER+ colon tumor, and was then primed with a Not I -  
oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
RI adaptors (Pharmacia), digested with Not I and cloned  
into the Not I and Eco RI sites of the modified p7T3  
vector. Library is normalized. Library was constructed by  
Bento Soares and M. Fatima Bonaldo (N-Soares4)."  
BASE COUNT 109 a 143 c 107 g 214 t  
ORIGIN  
Query Match 97.3%; Score 438; DB 36; Length 573;  
Best Local Similarity 100.0%; Pred. No. 9.3e-224;  
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 13 tgacccgagccatgacgagagcttgaggcgctgagagacagaggtggccgag 72  
Db 510 TGACCGCAGCCATGCGGACGAGGAGCTTGAGCGCTGAGGAGACAGAGGTCGCCGAG 451  
QY 73 ctgagggcaaacacaggggactctggtgatcgggcccaacaggaagacacagggaa 132  
Db 450 CTCGAGGCCAAACACAGGGGATCCTGGTGATCGGGCCCAACAGGAGCAAGCAGGGAA 391  
QY 133 gcagaaatgagaaacagtagtatcttagcccaagttctggatcagtcggccggccaggtta 192  
Db 390 GCAGAAATGAGAAACAGTATCTTAGCCCAAGTCTTGATCATCTCGCGCCGCCAGGTTA 331  
QY 193 agtaacttagcacttgaagcctgaaaaaactaaagcagtagagaattaccttatacag 252  
Db 330 AGTAACCTTAGCACTGTGAAGCCTGAAAAAAGCTAAAGCAGTAGAGAATTAACCTTATACAG 271  
QY 253 atggcaagatatggcaacactaagtgaaggtatcagaacaagggtttaataagaatcctt 312  
Db 270 ATGGCAAGATATGGCAACTAGTGAAGAGGTATCAGAACAAAGGTTTATAGAAATCTCT 211  
QY 313 aaaaagttaagcaaacacagaagacacaaacagtgaaattcaacagaagaagta 372  
Db 210 AAAAAGTAGTAAAGCCAAACAAAGAAAGAACACACAGTGAATTTCAACAGAAAGTA 151

QY 373 atggactctgatgaagatgacgattattgtaactacaaagtgctcacagactagaacttaac 432  
Db 150 ATGACTCTGATGAAGATGACGATTATTGTAAGTCTCAGACTAGAACTTAAC 91  
QY 433 ggaacaagtcttagacag 450  
Db 90 GGAACAAGTCTAGGACAG 73  
RESULT 3  
AI200921/c  
LOCUS AT200921 489 bp mRNA EST 14-OCT-1998  
DEFINITION qf63dl1.x1 Soares\_testis\_NHT Homo sapiens CDNA clone IMAGE:1754685  
3' similar to TR:O14737 O14737 TFAR19. ; mRNA sequence.  
ACCESSION AI200921  
VERSION AI200921.1 GI:3753527  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE 1 (bases 1 to 489)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 14, 1998 this sequence version replaced gi:1877660.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -40UP from GIBCO  
High quality sequence stop: 409.  
FEATURES  
source  
1. .489  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1754685"  
/clone\_lib="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/note="vector: p7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from Clontech  
Laboratories, Inc., and primed with a Not I - oligo(dT)  
primer [5].  
TGTTACCAATCTGAAGTGGGCGCGCCCAATTTTTTTTTTTT 3'.  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified p7T3 vector. Library  
went through one round of normalization to Cot5, and was  
constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 103 a 111 c 84 g 191 t  
ORIGIN  
Query Match 91.3%; Score 411; DB 43; Length 489;  
Best Local Similarity 100.0%; Pred. No. 2.6e-209;  
Matches 411; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 40 cttgaggcgctgagagacagagctggccagctgcaggccaaacacgggagctcgtt 99  
Db 489 CTTGAGCGCTGAGGAGACAGAGGCTGGCCGAGCTGCAGGCCAAACACGGGGATCCTGTT 430  
QY 100 gatcgcgcccaacaggaagcaaacagcaggggaagcagagaatgagaaacagtattcttagcc 159

```

|||||
Db 429 GATGCGCCCAACAGGAAGCAACAGCAGGGAAGCAAGAAATGAGAACAGTATCTTAGCC 370
QY 160 caagtctgatcagtcgccccgagcaggttaagtaacttagcacttgaagcctgaa 219
|||||
Db 369 CAAGTCTTGATCAGTCGGCCCGGCGAGGTTAAGTAACCTAGCACTTGTAAAGCCTGAA 310
QY 220 aaaaataaagcagtagagaattacaccttatacagatggcagaatattgacaactaagttag 279
|||||
Db 309 AAAACATAACAGCAGTACAGATTAACCTATACAGATGCGAAGATATGGACAACATAAGTGAG 250
QY 280 aaggtatcagaacaaggtttaaataagaaatccttaaaaaagtaagcgaacaaacagaaaaag 339
Db 249 AAGGTATCAGAACAAAGGTTTAAATAGAAATCCTTAAAAAAGTAAGCCCAACAAACAGAAAAG 190
QY 340 acacaacagtgaaatcaacagaagaagaaagtaagttagcctctgatgaagatgacgattat 399
Db 189 ACAACAACAGTGAATTCACACAGAGAAAGTAATGGACTCTGATGAGATGACGATTAT 130
QY 400 tgaactacaagtgctcacagactagaaacttaacgggaacaaagtgtaggacag 450
Db 129 TGAATACAAAGTGCTCACAGACTAGAACTTAACGGAACAAGTCTAGGACAG 79

RESULT 4
AA314244 408 bp mRNA EST 19-APR-1997
LOCUS EST186157 Colon carcinoma (HCC) cell line II Homo sapiens cDNA 5'
DEFINITION end, mRNA sequence.
ACCESSION AA314244
VERSION AA314244.1 GI:1966645
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 408)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitchugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glocke,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., HungJun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
TITLE Nature 377 (6547 Suppl), 3-174 (1995)
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 12140200
COMMENT On May 8, 1995 this sequence version replaced gi:801521.
Other_ESTs: THC169754
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavet@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
Location/Qualifiers
1. .408

```

```

/organism="Homo sapiens"
/db_xref="ATCC (inhost):110913"
/db_xref="taxon:9606"
/clone.lib="Colon carcinoma (HCC) cell line II"
/tissue_type="colon"
/cell_type="KM12C"
/cell_line="KM12C(HCC)-parental human colon
carcinoma; Dukes B2"
/Note="Organ: colon; Vector: pBluescript SK-; Site_1:
ECORI; Site_2: XhoI"
BASE COUNT 151 a 81 c 108 g 68 t
ORIGIN
Query Match 90.7%; Score 408; DB 31; Length 408;
Best Local Similarity 100.0%; Pred. No. 1.1e-207;
Matches 408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctgctccagcgtgacgcccagccatgcccagcagcgtgagcgtgagcgtgagcagcag 60
Db 1 CTGCTCCAGCGCTGACGCCGAGCCATGGCGGACGAGAGCTTGAGCGCTGAGGACAG 60
QY 61 aggcctggccgagcgtgcagcgaacacacggggtcctggtgagcgcgcacacaggaagca 120
Db 61 AGGCTGGCCGAGCTGCAGGCCCAACACACGGGGATCCTGTGTGATGGGCCCAACAGGAGCA 120
QY 121 aagcacaggggaagcagaataatgagaacagtagtctttagcccaagttcttgatcagtcggcc 180
Db 121 AAGCACAGGGAAGCAGAAAATGAGAAACAGTATCTTAGCCCAAGTTCTGGATCAGTCGGCC 180
QY 181 cgggcccaggttaagtaacttagcactgttaagcctgaaagcctgaaaaaactaaagcagtagaat 240
Db 181 CGGGCCAGGTTAAGTAACCTTAGCCTTGTAAAGCCCTGAAACCACTAAAGCAGTAGAAT 240
QY 241 taccttatacagatggcagaatgataagcactaagtgagaaggtatcacagaacaggttta 300
Db 241 TACCTTATACAGATGCGAAGATATGACAACTAAGTGAGAAGGTATCAGAAACAGGTTTA 300
QY 301 atagaataccttataaaaaagtaagcccaacaaacagaaaaagcaacacagtgaaattcaac 360
Db 301 ATAGAATACTTAAAAAAGTAAGCCCAACAAACAGAAAAGCAACACACAGTAAATTCAC 360
QY 361 agaagaaaagtaagtagcactctgataagatgacgattattgaactaca 408
Db 361 AGAAGAAAAGTAAATGGACTCTGTATGAAGATGACGATTATTGAACACTACA 408

RESULT 5
AA643309/c 468 bp mRNA EST 18-FEB-1998
LOCUS nr59b05.s1 NCI_CGAP_Lym3 Homo sapiens cDNA clone IMAGE:117241 3',
DEFINITION mRNA sequence.
ACCESSION AA643309
VERSION AA643309.1 GI:2568527
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 468)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor gene index
Unpublished (1997)
JOURNAL On Sep 12, 1996 this sequence version replaced gi:1405142.
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
WARNING: There is evidence that suggests that the 384-well parent
plate of this clone contains both human and mouse derived clones.
Thus, the origin of this clone is uncertain. This caution should be
kept in mind should you use this clone.

```

RESULT	6
AA416757	
LOCUS	
DEFINITION	AA416757 392 bp mRNA EST 16-OCT-1997 zu09a01.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:731304
ACCESSION	3', mRNA sequence.
VERSION	AA416757
KEYWORDS	AA416757.1 GI:2077711
SOURCE	EST.
ORGANISM	human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

|||||
Db 361 TTATGACTACAGTCTCAGACTAGAAC 392

RESULT 7
AA991276/c 500 bp mRNA EST 03-JUN-1998
LOCUS
DEFINITION
os51401.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1608840 3'
similar to TR:O14737 O14737 TFAR19. ;, mRNA sequence.
ACCESSION
AA991276
VERSION
AA991276.1 GI:3177765
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 500)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On Jan 9, 1998 this sequence version replaced gi:693448.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 368.
FEATURES
source
1..500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1608840"
/clone_lib="NCI_CGAP_Br2"
/sex="female, pooled"
/tissue_type="breast"
/lab_host="DH10B"
/notes="vector: p7T73-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from pooled bulk
breast tumor tissue, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified p7T73
vector. This library is the normalized version of
NCI_CGAP_Br1.1. Library was constructed by Bento Soares
and M. Fatima Bonaldo."
BASE COUNT 104 a 118 c 86 g 192 t
ORIGIN

Query Match 82.0%; Score 369; DB 40; Length 500;
Best Local Similarity 100.0%; Pred. No. 7.7e-187;
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 atgcgagcaggagcttgagcgctgaggagacagagctggcgagctcagggcaaaa 84
|||||
Db 500 ATGGCGGACGAGGAGCTTGAGCGCGCTGAGGAGACAGAGCTGGCGGAGCTCGAGGCCAAA 441
|||||
QY 85 caccggggatcctggtgtagcgcccaacaggaacaaacagcagggagcagaatgaga 144
|||||
Db 440 CACGGGATCTGTGTATGCGGCCCCACAGAACGACGAGGACAGGAGCAATGAGA 381
|||||
QY 145 aacagtatcttagcccaagttctgtagtcagtcggcccgggccaggttaagttaattagca 204
|||||
Db 380 AACAGTATCTTAGCCCAAGTTCTGATCAGTCGCGCCGCGGCGCAGGTTAAGTAACTTAGCA 321
|||||

```

```

QY 205 cttgtaaaagcctgaaaaaactaaagcagtagagaattacottatcacagatggcaagatat 264
|||||
Db 320 CTTGTTAAAGCCTGAAAAAACTAAAGCAGTAGAGATTACCTTTATACAGATGGCAAGATAT 261
|||||
QY 265 ggacaactaagtgaaggttatcagacaaggttttaataagaaatccttaaaaaagtaagc 324
|||||
Db 260 GGACAACTAAGTCAGAGAGGTATCAGAACAGGGTTTATAGAAATCCTTAAAAAAGTAAGC 201
|||||
QY 325 caacaacagaaaaagacaacacacagtgaaattccaacagagaagaaagtaagtgcctctgat 384
|||||
Db 200 CAACAAACAGAAAAAGACACACACAGTGAAATTCAACAGAGAAAAAGTAATGGACTCTGAT 141
|||||
QY 385 gaagatgac 393
|||||
Db 140 GAAGATGAC 132
|||||

RESULT 8
AA125964/c 498 bp mRNA EST 27-OCT-1998
LOCUS
DEFINITION
qc46b07.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
IMAGE:1712629 3', similar to TR:O14737 O14737 TFAR19. ;, mRNA
sequence.
ACCESSION
AA125964
VERSION
AA125964.1 GI:3594478
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 498)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On Jan 17, 1998 this sequence version replaced gi:2044790.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 607 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 394.
FEATURES
source
1..498
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1712629"
/clone_lib="Soares_pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Organ: uterus; Vector: p7T73-Pac; Site.1: Not I -
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
AACTGGAGAAATTCGGCGCCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
BASE COUNT 104 a 117 c 86 g 191 t
ORIGIN

Query Match 80.7%; Score 363; DB 42; Length 498;
Best Local Similarity 100.0%; Pred. No. 1.2e-183;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ggggatacttggtgatcgcccaacaggaagcaagcacagggaagcagaatgagaac 147
|||||

```









DEFINITION ge09g02.x1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:1738418  
3' similar to TR:014737 O14737 TFAR19. ; mRNA sequence.

ACCESSION AI187865  
VERSION AI187865.1 GI:3739074  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 500)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)

Insert Length: 596 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 317.  
Location/Qualifiers  
1..500  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="1738418"  
/clone\_lib="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from Clontech  
Laboratories, Inc., and primed with a Not I - oligo(dT)  
primer [5'  
TGTTACCAATCGAAGTGGAGCGGCCCAATTTTTTTTTTTT 3'].  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization to Cot5, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 104 a 114 c 87 g 191 t 4 others  
ORIGIN

Query Match 67.6%; Score 304; DB 43; Length 500;  
Best Local Similarity 100.0%; Pred. No. 4.5e-152;  
Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 cagtattctagccaggttctgattcagtcagtcgcccggccaggttaagtaacttagcact 206  
|||||  
Db 381 CAGTATCTTAGCCCAAGTCTGTGATCAGTCAGTCGCCGCCGAGTTAAGTAACTTAGCACT 322  
|||||  
QY 207 tgtaaagcctgaaataactaaacagtagagaattacccttatacagatggcaagatatgg 266  
|||||  
Db 321 TGTAAGCCTGAAATAACTAAACAGTAGAGATTAACCTTATACAGATGGCAAGATATGG 262  
|||||  
QY 267 acaactaagtgaaggtatcagaacaaggtttaataagaataccttaaaaaagtaagcca 326  
|||||  
Db 261 ACAACTAAGTGAGAAGGTATCAGAACAAAGGTTTAATAGAAATCCTTAAAAAAGTAAGCCA 202  
|||||  
QY 327 acaacagaaacacacaacagtcgaattcaacagagaagaagaatgaactctgatga 386  
|||||  
Db 201 ACAACAGAAAGACACACACAGTGAATTCACAGAGAAAGTAATGGACACTCTGATGA 142  
|||||  
QY 387 agatgacgattattgaactacaagtgctcacagactagaacttaacggaacaagtcctagg 446  
|||||

Db 141 AGATGACGATTATTGAACATAAGTCTCACAGACTAGAACTTAACGGAAACAGTCTAGG 82

QY 447 acag 450  
|||||  
Db 81 ACAG 78

RESULT 15  
LOCUS R81019 426 bp mRNA EST 09-JUN-1995  
DEFINITION Y194g04.r1 Soares placenta Nb2HP Homo sapiens cDNA clone  
IMAGE:146934 5', mRNA sequence.  
ACCESSION R81019  
VERSION R81019.1 GI:857300  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 426)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT On May 8, 1995 this sequence version replaced gi:801702.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
Insert Size: 608  
High quality sequence stops: 309  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 608 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 309.  
Location/Qualifiers  
1..426  
/organism="Homo sapiens"  
/db\_xref="GDB:558531"  
/db\_xref="taxon:9606"  
/clone\_image="146934"  
/clone\_lib="Soares placenta Nb2HP"  
/sex="female"  
/dev\_stage="placenta obtained at birth (full term)"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
AAGTGAAGATTCGCGCCGAGAAATTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 155 a 76 c 109 g 81 t 5 others  
ORIGIN

Query Match 62.9%; Score 283; DB 22; Length 426;  
Best Local Similarity 100.0%; Pred. No. 7.8e-141;  
Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 agctgcagcccaaacacggggatccttggtgtagcgcccaacaggaagcaagcacaggg 130  
|||||  
Db 26 AGCTGCAGCCAAACACGGGGATCCTGTTGATGTCGCCGCCCAACAGGAAGCAACAGCAGGG 85

Qy 131 aagcagaaatgagaaacagtgatcttagcccaagttcttgatcgccggccaggt 190  
|||  
Db 86 AAGCAGAAATGAGAAACAGTATCTTAGCCCAAGTTCTGATCAGTCGCCGGGCCAGGT 145  
Qy 191 taagtaacttagcacttgtaagccctgaaaaaactaaagcagtagagaaattaccttatac 250  
|||  
Db 146 TAAGTAACTTAGCCTTGTAAAGCCTGAAAAAATTAAGCAGTAGAGAAATTACCTTATAC 205  
Qy 251 agatggcaagatatggacaactaagtgaaggtatcagaacaagtttaataagaaatcc 310  
|||  
Db 206 AGATGGCAAGATATGGACAACCTAAGTGAGAGGTATCAGAACAGGTTTAATAGAAATCC 265  
Qy 311 ttaaaaaagtaagcccaacacagaaaaagacaacacagtgaa 353  
|||  
Db 266 TTAAAAAAGTAGCCCAACAAACAGAAAAAGACAACACAGTGAA 308

Search completed: May 1, 2000, 14:44:37  
Job time: 13997 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:32:40 ; Search time 4425.31 Seconds  
(without alignments)  
-804.843 Million cell updates/sec

Title: US-09-215-435-117  
Perfect score: 1173  
Sequence: 1 gagctgttatggacacg.....tccatctcaaaaaaaaaa 1173

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0  
Searched: 821193 seqs, -1518192014 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_on.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_p11.\*
- 8: gb\_p12.\*
- 9: gb\_pri.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_v1.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_on.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sy.\*
- 29: em\_un.\*
- 30: em\_v1.\*
- 31: em\_v2.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_ba1.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*

- 44: gb\_htg6.\*
- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	23	2.0	123339	41	AF179296 Homo sapi
C 2	22	1.9	1381	35	AF016835 Drosophil
C 3	22	1.9	4132	40	MI3918 Homo sapien
C 4	22	1.9	38875	40	AC006950 Homo sapi
C 5	22	1.9	39519	42	AC012054 Homo sapi
C 6	22	1.9	42664	41	AC011559 Homo sapi
C 7	22	1.9	74523	44	AC016158 Drosophil
C 8	22	1.9	109725	44	AC010199 Homo sapi
C 9	22	1.9	120879	41	AC008364 Drosophil
C 10	22	1.9	127207	33	AC007991 Homo sapi
C 11	22	1.9	132398	44	AC010193 Homo sapi
C 12	22	1.9	134450	11	HUAC003049
C 13	22	1.9	135505	11	AC004477 Homo sapi
C 14	22	1.9	139149	44	AC011163 Homo sapi
C 15	22	1.9	139180	41	AC007712 Drosophil
C 16	22	1.9	142228	44	AC004123 Homo sapi
C 17	22	1.9	153400	42	AC009700 Homo sapi
C 18	22	1.9	154959	40	AC004925 Homo sapi
C 19	22	1.9	155407	11	AF042090 Homo sapi
C 20	22	1.9	156192	32	HSBA775A3
C 21	22	1.9	166663	33	AC007728 Homo sapi
C 22	22	1.9	183570	41	AC007906 Homo sapi
C 23	22	1.9	186323	44	AC006491 Drosophil
C 24	22	1.9	186395	33	HS393J16
C 25	22	1.9	187820	33	AC007222 Homo sapi
C 26	22	1.9	195898	41	AC011479 Homo sapi
C 27	22	1.9	204585	32	HSJ324N14
C 28	22	1.9	205784	41	AC008367 Drosophil
C 29	22	1.9	250687	43	AC011601 Homo sapi
C 30	22	1.9	312283	40	AF205588 Homo sapi
C 31	21	1.8	302	9	HUMRSAB2
C 32	21	1.8	492	9	HSALPS
C 33	21	1.8	551	13	G50756
C 34	21	1.8	563	10	PTU38668
C 35	21	1.8	1319	12	MUSCHGA
C 36	21	1.8	1488	9	HS393J16
C 37	21	1.8	1664	5	AF014240
C 38	21	1.8	1664	5	I73181
C 39	21	1.8	1665	9	HUMASMA
C 40	21	1.8	1750	1	STMAFSB
C 41	21	1.8	1946	10	HUMG1A10
C 42	21	1.8	2060	10	HS393J16
C 43	21	1.8	2115	5	E00256
C 44	21	1.8	2287	9	HUMNCAW
C 45	21	1.8	2287	9	HUMNCAW

ALIGNMENTS

RESULT 1  
AF179296/c  
LOCUS AF179296 123339 bp DNA 08-OCT-1999  
DEFINITION Homo sapiens, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
ACCESSION AF179296  
VERSION AF179296.1 GI:5802962

```

KEYWORDS      HTG; HTGS_PHASE2.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS        Eutheria; Primates; Catarrhini; Hominidae; Homo.
               1 (bases 1 to 123339)
               Schudy,A., Platzer,M., Schlhabel,M., Koczan,D., Thiesen,H.-J.,
               Merck-Rousseau,M.F., Baumgart,C., Menzel,U., Weber,J.,
               Schattevoy,R. and Rosenthal,A.
TITLE          Direct Submission
JOURNAL        Submitted (20-AUG-1999) Genome Analysis, Institute of Molecular
COMMENT        Biotechnology, Butenbergsstrasse 11, Jena 07745, Germany
               contig 1: pos. 1 - 92151 contig 2: pos. 92152 - 123339 Inbetween a
               gap of unknown size.
               * NOTE: This is a 'working draft' sequence.
               * This sequence will be replaced
               * by the finished sequence as soon as it is available and
               * the accession number will be preserved.

FEATURES       Location/Qualifiers
               1..123339
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               complement(2..144)
               /rpt_family="MIR"
               395..438
               /note="SST_REPEAT"
               complement(556..676)
               /rpt_family="MIR"
               566..636
               /rpt_family="mir"
               complement(746..930)
               /rpt_family="MIR"
               1036..1142
               /rpt_family="alu"
               1045..1119
               /rpt_family="sva"
               1055..1182
               /rpt_family="alu"
               1055..1180
               /rpt_family="ALU"
               1055..1335
               /rpt_type="inverted"
               1055..1331
               /rpt_family="AluX"
               complement(1077..1179)
               /rpt_family="SVA"
               1175..1232
               /rpt_family="sva"
               1175..1232
               /rpt_family="alu"
               1229..1333
               /rpt_family="alu"
               1232..1335
               /rpt_family="ALU"
               1272..1346
               /note="SST_REPEAT"
               1272..1346
               /note="XNUN_REPEAT"
               complement(1332..1628)
               /rpt_family="AluX"
               complement(1339..1627)
               /rpt_family="alu"
               complement(1339..1628)
               /rpt_family="ALU"
               1347..1628
               /rpt_type="inverted"
               complement(1366..1423)
               /rpt_family="sva"
               complement(1410..1474)
               /rpt_family="sva"
               complement(1412..1488)
               /rpt_family="alu"
               complement(1431..1515)

               repeat_region      /rpt_family="sva"
               complement(1502..1605)
               /rpt_family="sva"
               1506..1605
               /rpt_family="SVA"
               complement(1535..1628)
               /rpt_family="alu"
               complement(1566..1628)
               /rpt_family="sva"
               1624..1730
               /rpt_family="MIR"
               complement(1624..1737)
               /rpt_family="mir"
               1629..1797
               /rpt_family="MIR"
               1796..1841
               /note="XNUN_REPEAT"
               1799..1841
               /note="SST_REPEAT"
               complement(1800..1985)
               /note="GRAIL"
               /evidence=not_experimental
               1966..2073
               /note="SST_REPEAT"
               2756..2847
               /rpt_family="alu"
               2756..2824
               /rpt_family="sva"
               2758..3051
               /rpt_family="AluX"
               2758..2869
               /rpt_family="alu"
               2777..2872
               /rpt_family="sva"
               complement(2777..2868)
               /rpt_family="SVA"
               complement(2822..3037)
               /note="FEXHB"
               /evidence=not_experimental
               2863..3047
               /rpt_family="ALU"
               2867..3048
               /rpt_family="alu"
               2883..2954
               /rpt_family="sva"
               2894..2987
               /rpt_family="alu"
               2913..2976
               /rpt_family="sva"
               2962..3017
               /rpt_family="sva"
               3041..3105
               /note="SST_REPEAT"
               3057..3208
               /rpt_family="l1mc3"
               complement(3070..3271)
               /rpt_family="L1MC3"
               3137..3205
               /rpt_family="l1mc4"
               3138..3205
               /rpt_family="l1mc2"
               complement(3142..3207)
               /rpt_family="L1MC3"
               complement(3371..3421)
               /rpt_family="mir"
               3372..3545
               /rpt_family="MIR"
               complement(3405..3461)
               /rpt_family="mir"
               complement(3721..3802)
               /rpt_family="mir"
               3762..3900
               /rpt_family="MIR"

```

```

repeat_region 4007. .4055
/note="SST_REPEAT"
repeat_region 4108. .4318
/rpt_family="MSTB"
repeat_region 4108. .4246
/rpt_family="MSTA"
repeat_region complement(4110. .4245)
/rpt_family="msta"
repeat_region complement(4154. .4218)
/rpt_family="thelb"
exon 4211. .4246
/note="XPOUND"
/evidence-not_experimental
repeat_region 4290. .4414
/rpt_family="alu"
repeat_region 4319. .4619
/rpt_family="AluSq"
repeat_region 4322. .4449
/rpt_family="sva"
repeat_region 4322. .4491
/rpt_family="alu"
repeat_region 4332. .4484
/rpt_family="ALU"
repeat_region complement(4342. .4443)
/rpt_family="SVA"
repeat_region 4345. .4706
/rpt_type="inverted"
repeat_region 4347. .4706
/rpt_type="inverted"
repeat_region 4349. .4706
/rpt_type="inverted"
repeat_region 4351. .4417
/rpt_type="inverted"
repeat_region 4465. .4606
/rpt_family="alu"
repeat_region 4524. .4579
/rpt_family="sva"
repeat_region 4709. .5070
/rpt_type="inverted"
repeat_region 4709. .5066
/rpt_type="inverted"
repeat_region 4709. .5068
/rpt_type="inverted"
repeat_region 4775. .5004
/note="XNUN_REPEAT"
repeat_region 4775. .4866
/rpt_type="tandem"
repeat_region complement(4775. .4866)
/note="(TGA)n"
/rpt_family="Simple_repeat"

Query Match 2.0%; Score 23; DB 41; Length 123339;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 acgtgtgtgtgcgcgcgcacc 62
|||||
Db 55970 ACGTGTGTGTGCGCGCCACC 55948

RESULT 2
AF016835 1381 bp mRNA INV 14-JUL-1998
LOCUS Drosophila melanogaster ribosomal protein L3 (RpL3) mRNA, complete
DEFINITION cds.
ACCESSION AF016835
VERSION AF016835.1 GI:2384753
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

```

```

REFERENCE 1 (bases 1 to 1381)
AUTHORS Chan,H.Y.E., Zhang,Y., Hohelsel,J.D. and O'Kane,C.J.
TITLE Identification and characterization of the gene for Drosophila L3
ribosomal protein
JOURNAL Gene 212 (1), 119-125 (1998)
MEDLINE 98326317
REFERENCE 2 (bases 1 to 1381)
AUTHORS Chan,H.Y.E., Zhang,Y., Hohelsel,J.D. and O'Kane,C.J.
TITLE Direct Submission
JOURNAL Submitted (31-JUL-1997) Genetics, University of Cambridge, Downing
Street, Cambridge CB2 3EH, England
FEATURES
source
1. .1381
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3"
/map="86D"
1. .1381
/gene="RpL3"
34. .1284
/gene="RpL3"
/note="ribosomal protein"
/codon_start=1
/product="ribosomal protein L3"
/protein_id="AAC26144.1"
/db_xref="GI:2384754"
/translation="MSHRKFSAPRHGSMFYPKRSARHGKVKAFPPKDDASKPVHLT
CFIYKAGTHIVREADRPGSKNKKEVEAVTVLETPMIVVGAIVYETPFLRAL
VNWQHLSEECRRFRYKNYKSKKAFKASKWTDDLGKSIENFRKMLRYCKVI
RVIAHSQJELIKORQKKAHVMEIQLNGSIEDKVKWAREHLEKPIQVSNVFGQDEMID
RVEINKKIYRIGAGIHTKDGKVIKNNASTYDLTKSITPMGGFPHYGVNDFVMIK
GCCIGSKRIITLRKSLKHTKRSALQIKLFIDTSSKMGHGRFQTPADKLAFMGPL
KKDLRKEFAAATAAAAATTSA"
BASE COUNT 324 a 421 c 373 g 263 t
ORIGIN

Query Match 1.9%; Score 22; DB 35; Length 1381;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 cgtgtgtgtgcgcgcgcacc 62
|||||
Db 1248 CGTGTGTGTGCGCGCCACC 1269

RESULT 3
HUMFNRRAS 4132 bp mRNA PRI 22-MAR-1999
LOCUS Homo sapiens fibronectin receptor alpha-subunit precursor (ITGA5)
DEFINITION mRNA, partial cds.
ACCESSION M13918 M18364
VERSION M13918.2 GI:4464190
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 2416 to 4132)
AUTHORS Argaves,W.S., Pytela,R., Suzuki,S., Millan,J.L.,
Pierschbacher,M.D. and Ruoslahti,E.
TITLE cDNA sequences from the alpha subunit of the fibronectin receptor
predict a transmembrane domain and a short cytoplasmic peptide
JOURNAL J. Biol. Chem. 261 (28), 12922-12924 (1986)
MEDLINE 87008489
REFERENCE 2 (bases 1 to 3281)
AUTHORS Fitzgerald,L.A., Poncz,M., Steiner,B., Rall,S.C. Jr., Bennett,J.S.
and Phillips,D.R.
TITLE Comparison of cDNA-derived protein sequences of the human
fibronectin and vitronectin receptor alpha-subunits and platelet
glycoprotein IIb
JOURNAL Biochemistry 26 (25), 8158-8165 (1987)

```

AUTHORS	Lamerdin,J.E., McCready,P.M., Skowronski,E., Wiswanathan,V., Burkhart-Schultz,K., Gordon,L., Dias,J.J., Ramirez,M., Stillwagen,S., Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J., Danganan,L., Erler,A., Christensen,M., Georgescu,A., Avila,J., Liu,S., Attix,C., Andreise,T., Trankhelm,M., Amico-Keller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G., Kronmiller,B., Arellano,A., Sanders,C., Ow/D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
TITLE	Sequence analysis of a 2.2 Mb region in 19q13.1 containing the RYR gene
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 38875)
AUTHORS	Lamerdin,J.E.
TITLE	Direct Submission
JOURNAL	Submitted (03-MAR-1999) Joint Genome Institute, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT	Map and sequence oriented from q centromere to telomere. Cosmid R30669 overlaps BAC 331004 (CIR-B-47On8; AC005393) to the left from bases 1 to 8,001 of this accession, and overlaps cosmid R28471 to the right from bases 31,491 to 38,875. For updates to our chromosome 19 map and sequence information, visit our web site at: http://www.bio.lnl.gov/bbrp/genome/genome.html.
FEATURES	Location/Qualifiers 1..38875 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="R30669" /chromosomes="19" /map="19q13.1 between D19S220 and AKT2" /cell_line="5HL2-B" /clone_lib="LL19NC03 R chromosome 19-specific cosmid library" /note="Cosmid library constructed at LLNL from flow-sorted chromosomes from hybrid 5HL2-B, which carries chromosome 19 as its only human chromosome." repeat_region complement(178..303) /rpt_family="LINE2" mRNA complement(join<1003..1134,2326..2496,3524..3618,3705..3806,3908..4078,9582..9618)) /gene="FBL" gene /product="Human fibrillarlin (Hfibl) mRNA, partial cds" complement(<1003..9618) /gene="FBL" /note="fibrillarlin" cbs complement(join<1003..1134,2326..2496,3524..3618,3705..3806,3908..4078,9582..9591)) /gene="FBL" /function="NUCLEAR PROTEIN" /function="RNA PROCESSING" /function="RNA-BINDING" /function="RIBONUCLEOPROTEIN" /note="FIBRILLARIN; 34 KD NUCLEAR SCLERODERMA ANTIGEN" /codon_start=1 /product="FBL_HUMAN [AA 1- 227]" protein_id="AADI5623.1" db_xref="GI_4321126" translation=MRPGSPRGGGGGRGCGFGDRGGRGCGFGGGRGGGFRGG RGGGGGGGGGGGGGGSGNRRGGRGGRGNQSNWVEPHRHGVICRKG EDALVTKLVPSVTGKEKRVSTSEGDDKIETRWPNFRSKLAAILGGVDYQHIKP AKHYLCGAAGTGTVSHVSIVDPDGLVAVFEFSHRSGRDILNAKRNTIIPVEDIA HPKVRLI repeat_region complement(1136..1156) /rpt_family="At_rich" repeat_region complement(1209..1341) /rpt_family="MIR" repeat_region complement(1428..1539) /rpt_family="MIR" repeat_region complement(1556..1851) /rpt_family="AlusX" repeat_region complement(2723..3090) /rpt_family="LINE2" repeat_region complement(3219..3430) /rpt_family="LINE2"

```

MEDLINE      88163472
COMMENT      On Mar 22, 1999 this sequence version replaced gi.182709..
              Draft entry and sequence in computer-readable form for [1] kindly
              provided by W.S.Argraives, 08-DEC-1986.

FEATURES
source       1..4132
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /map="12q11-q13"
             /chromosome="12"
             /tissue_type="placenta; umbilical vein endothelium"
             <1..4132
             /gene="ITGA5"
             /note="FNRA"
             <1..3105
             /gene="ITGA5"
             /note="surface glycoprotein receptor"
             /codon_start=1
             /product="fibronectin receptor alpha-subunit precursor"
             /protein_id="AAAS2467.1"
             /db_xref="GI:182710"
             /translation="RWGFRPPPLPLLLLLPPPRVGGFNDAEAPAVLSGPPGSG
             FGSVEFRPGTDGVSILVIGAPKANTSQVLQGGAVLCPMGASPTQCTPIEFDSKG
             SRLLESLSSEEGEPVEYKSLQWFGATVRAHGSSILACAPLYSWRTEKEPLSDPVG
             CYLSDNFTFRIELAPCRSDFSWAAGQGVCOGGFAETKTGRVVLGGPGSYFWQGI
             LSATQEDFAEYSPVLLNLVQGLQTRQASSIYDYSVLGYSVAGFSGDDTDFVA
             GVPKNLYGYVTILNGSDIRSLYNFSEQMASYEGYAAATDVNGDGLDLDLLVGAPL
             LMDRTDPGRQEVGRVYTLQHPAGIEPTLTLLGHDEFGRGSSLPTPLGDLDDQGY
             NDVAIGAPGGTQGVVEVFFGGGGLGSKPSQVLQPLAAASHIPDFFGSALRGGR
             LDNGNYCLDILVSGFDKAVYRGRIIVSASASLTFFPAMFPEERSCSLEGNPVACI
             NLSFCLNAGIKVADSIQFTVELQMDWKQGVRRALFLASTQATLTQTLLIQNGAR
             EDCREKTYILRNESFDRKLSPIHIALNFSLDQPVDSHGLRPALHYQSKSRIEDTA
             QIILDCGGEINCPVDLQEVGEQNHVYLGDKNALNLTPHAONVGEGAYEAELAVTA
             PPEAEGSLVRPFGNFSSISCDYFAVNQSRLLVCDLGNPKAGASLWGLRFTVPLRLR
             DTKKTIQDFQILSKNLNNSQDVFSRLVSEAQVTLNGVSKPEAVLFPVSDNHPR
             DQPOKEEDGPVAVHHYIELINQGPSISQVLELSCPOALEQQLLYIVRVGLNCTT
             NHPINPKLEDPESGLHHQKREAPSRSSASGPQILKCAPEACFRLCRCEGLPHLHQ
             ESQSLQHLFRVAKVFLQREHGFSLQCEAVYKALKMPYRIPLQPLQKRGQVATAVQ
             WTKAEGSGVPLWIIILAILFGLLGLLIYILKLGPFKRSILPYGTAMEKAQLKAPPA
             TSDA"
sig_peptide <1..78
             /gene="ITGA5"
mat_peptide 79..2652
             /gene="ITGA5"
             /product="fibronectin receptor alpha-subunit large chain"
mat_peptide 2653..3102
             /gene="ITGA5"
             /product="fibronectin receptor alpha-subunit small chain"

BASE COUNT 903 a 1261 c 1125 g 843 t

Query Match      1.9%; Score 22; DB 40; Length 4132;
Best Local Similarity 100.0%; Pred. No. 0.99;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 42 gctgctgctgcgcgcgcaccc 63
      |||
Db 45 GCTGCTGCTGCGCGCCACCC 66

RESULT 4
LOCUS AC006950 38875 bp DNA PRI 03-MAR-1999
DEFINITION Homo sapiens chromosome 19, cosmid R30669, complete sequence.
ACCESSION AC006950
VERSION AC006950.1 GI:4321125
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Primates; Catarrhini; Hominidae; Homo.
FEATURES
          1..bases 1 to 38875

```



```
repeat_region complement(3907..4016)
repeat_region /rpt_family="(GGA)n"
repeat_region 4116..4273
repeat_region /rpt_family="LIME"
repeat_region complement(4356..4646)
repeat_region /rpt_family="AluJo"
repeat_region complement(5377..5676)
repeat_region /rpt_family="AluY"
repeat_region 5861..6011
repeat_region /rpt_family="LINE2"
repeat_region complement(6042..6126)
repeat_region /rpt_family="MER5B"
repeat_region 6179..6545
repeat_region /rpt_family="LINE2"
repeat_region 6779..7078
repeat_region /rpt_family="AluSx"
repeat_region 7775..8156
repeat_region /rpt_family="LINE2"
repeat_region complement(8685..8765)
repeat_region /rpt_family="MIR"
repeat_region complement(10030..10749)
repeat_region /rpt_family="LINE2"
repeat_region 10887..10976
repeat_region /rpt_family="LINE2"
repeat_region 11106..11402
repeat_region /rpt_family="AluJo"
repeat_region complement(11730..11805)
repeat_region /rpt_family="LINE2"
repeat_region 12330..13004
repeat_region /rpt_family="LINE2"
repeat_region 13043..13188
repeat_region /rpt_family="MLT1A2"
repeat_region complement(13189..13490)
repeat_region /rpt_family="AluSp"
repeat_region 13528..13678
repeat_region /rpt_family="MLT1A2"
repeat_region complement(13692..13740)
repeat_region /rpt_family="LMB7"
repeat_region 13742..14034
repeat_region /rpt_family="AluSc"
repeat_region complement(14035..14130)
repeat_region /rpt_family="LMB8"
repeat_region complement(14130..14469)
repeat_region /rpt_family="LMB8"
repeat_region complement(14480..14781)
repeat_region /rpt_family="AluJo"
repeat_region 14803..15004
repeat_region /rpt_family="LINE2"
repeat_region complement(15068..15155)
repeat_region /rpt_family="LIPAL5"
repeat_region complement(15162..15445)
repeat_region /rpt_family="AluSg"
repeat_region complement(15639..15724)
repeat_region /rpt_family="(TA)n"
repeat_region complement(15725..16021)
repeat_region /rpt_family="AluSg"
repeat_region complement(16145..16535)
repeat_region /rpt_family="MSTD"
repeat_region 16592..16832
repeat_region /rpt_family="AluY"
repeat_region 16833..16976
repeat_region /rpt_family="AluSg/x"
repeat_region 16991..17325
repeat_region /rpt_family="LIMC1"
repeat_region 17330..17641
repeat_region /rpt_family="AluJo"
repeat_region 17670..17835
repeat_region /rpt_family="FRAM"
repeat_region complement(17924..18232)
repeat_region /rpt_family="AluSx"
repeat_region 18243..18771
repeat_region /rpt_family="LINE2"
misc_feature complement(18325..18444)

/note="predicted exon, program: gail2exons_human_1.3,
frame: 1, quality: good, score: 62.000"
/rpt_family="AluSx"
complement(19138..19370)
/rpt_family="MLTIC"
19399..19610
/rpt_family="AluSg"
complement(19618..19804)
/rpt_family="MLTIC"
complement(19863..19980)
/rpt_family="FLAM_C"
complement(20138..20477)
/rpt_family="LIMC4"
20478..20782
/rpt_family="AluSg"
20804..21109
/rpt_family="AluSx"
complement(21224..21337)
/rpt_family="AluJo"
complement(21768..22315)
/rpt_family="MER51B"
22391..22686
/rpt_family="AluSx"
complement(22695..22891)
/rpt_family="LIMC/D"
complement(22927..23211)
/rpt_family="AluSx"
23383..23497
/rpt_family="MIR"
complement(23633..23937)
/rpt_family="AluSx"
complement(24057..24291)
/rpt_family="LINE2"
24536..24645
/rpt_family="(TA)n"
complement(24687..24926)
/rpt_family="AluJo"
complement(25086..25369)
/rpt_family="AluSg"
25466..25764
/rpt_family="AluSc"
complement(25907..26007)
/rpt_family="MIR"

Query Match 1.9%; Score 22; DB 40; Length 38875;
Best Local Similarity 100.0%; Pred. No. 0.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1152 cttccatctcaaaaaaaaaa 1173
|||||
DB 20753 CTTCCATCTCAAAAAAAAAA 20774

RESULT 5
AC012054/c 39519 bp DNA HTG 19-OCT-1999
LOCUS Homo sapiens chromosome 11 clone 90_G_01 map 11, LOW-PASS SEQUENCE
DEFINITION SAMPLING.
ACCESSION AC012054
VERSION AC012054.1 GI:6067189
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 39519)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone 90_G_01
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 39519)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
```

Brown, A., Castile, A., Colangelo, M., Collins, S., Collymore, A.,  
Cooke, P., DeArellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,  
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Horton, L.,  
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Klein, J.,  
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,  
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,  
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,  
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,  
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (19-OCT-1999), Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker: Smit, A. F. A. &  
Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.  
\* NOTE: This record contains 49 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

TITLE  
JOURNAL

COMMENT

17756	18570:	contig of 815 bp in length
*		gap of unknown length
*	18571	contig of 822 bp in length
*		gap of unknown length
*	19393	contig of 800 bp in length
*		gap of unknown length
*	20193	contig of 819 bp in length
*		gap of unknown length
*	21012	contig of 809 bp in length
*		gap of unknown length
*	21821	contig of 768 bp in length
*		gap of unknown length
*	22589	contig of 825 bp in length
*		gap of unknown length
*	23414	contig of 783 bp in length
*		gap of unknown length
*	24197	contig of 804 bp in length
*		gap of unknown length
*	25001	contig of 825 bp in length
*		gap of unknown length
*	25826	contig of 808 bp in length
*		gap of unknown length
*	26634	contig of 817 bp in length
*		gap of unknown length
*	27451	contig of 840 bp in length
*		gap of unknown length
*	28291	contig of 785 bp in length
*		gap of unknown length
*	29076	contig of 816 bp in length
*		gap of unknown length
*	29892	contig of 782 bp in length
*		gap of unknown length
*	30674	contig of 793 bp in length
*		gap of unknown length
*	31467	contig of 794 bp in length
*		gap of unknown length
*	32261	contig of 803 bp in length
*		gap of unknown length
*	33064	contig of 839 bp in length
*		gap of unknown length
*	33903	contig of 829 bp in length
*		gap of unknown length
*	34732	contig of 816 bp in length
*		gap of unknown length
*	35548	contig of 764 bp in length
*		gap of unknown length
*	36312	contig of 767 bp in length
*		gap of unknown length
*	37079	contig of 821 bp in length
*		gap of unknown length
*	37900	contig of 811 bp in length
*		gap of unknown length
*	38711	contig of 809 bp in length.

FEATURES  
Location/Qualifiers  
1..39519  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11"  
/clone="90\_G\_01"  
/clone\_lib="RPCI-11 Human Male BAC"

BASE COUNT 10815 a 8660 c 8902 g 10353 t 789 others  
ORIGIN

Query Match 1.9%: Score 22: DB 42: Length 39519;  
Best Local Similarity 100.0%: Pred. No. 0.69;  
Matches 22: Conservative 0: Mismatches 0: Indels 0: Gaps 0;

QY 1152 cttccatctcaaaaaaaaaa 1173  
|||||  
Db 6185 CTTCATCTCAAAAAAAAAA 6164

```
RESULT 6
AC011559
LOCUS
DEFINITION Homo sapiens chromosome 19 clone LLNL-R_300G1, HTG 07-OCT-1999
PROGRESS ***, 1 ordered pieces.
AC011559
VERSION AC011559.1 GI:6015193
KEYWORDS HTG; HTGS_PHASE2.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 42664)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 42664)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
www.jgi.doe.gov#mapping database ID 33777.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved
* 1 42664: contig of 42664 bp in length.
FEATURES
source
1..42664
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="LLNL-R_300G1"
BASE COUNT 10899 a 10101 c 10058 g 11606 t
ORIGIN

Query Match 1.9%; Score 22; DB 41; Length 42664;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1152 cttccatctcctcaaaaaaaaaa 1173
|||||
Db 33322 CTTCCATCTCAAAAAAAAAA 33343

RESULT 7
AC016158/c
LOCUS
DEFINITION Drosophila melanogaster chromosome 3 clone BACR38H07 (D1203)
RPII-98.38.H.7 map 92D-92D strain Y; cn bw sp, *** SEQUENCING IN
PROGRESS ***, 69 unordered pieces.
AC016158
VERSION AC016158.1 GI:6466505
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephyarotidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 74523)
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richardson,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Direct Submission
Submitted (23-NOV-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 69 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1317: contig of 1317 bp in length
1318 1397: gap of unknown length
1398 1995: contig of 598 bp in length
1996 2075: gap of unknown length
2076 2858: contig of 783 bp in length
2859 2938: gap of unknown length
2939 3636: contig of 698 bp in length
3637 3716: gap of unknown length
3717 5148: contig of 1431 bp in length
5148 5228: gap of unknown length
5228 5988: contig of 760 bp in length
5988 6067: gap of unknown length
6068 6769: contig of 702 bp in length
6770 6849: gap of unknown length
6850 7674: contig of 825 bp in length
7675 7754: gap of unknown length
7755 8519: contig of 765 bp in length
8520 8599: gap of unknown length
8600 9488: contig of 889 bp in length
9489 9569: gap of unknown length
9569 10567: contig of 999 bp in length
10568 10648: gap of unknown length
10648 12209: contig of 1562 bp in length
12210 12289: gap of unknown length
12290 13010: contig of 721 bp in length
13011 13090: gap of unknown length
13091 14499: contig of 1409 bp in length
14500 14580: gap of unknown length
14580 16194: contig of 1614 bp in length
16194 16274: gap of unknown length
16274 17532: contig of 1259 bp in length
17533 17613: gap of unknown length
17613 19098: contig of 1486 bp in length
19099 19178: gap of unknown length
19179 20704: contig of 1526 bp in length
20705 20785: gap of unknown length
20785 22171: contig of 1386 bp in length
22171 22551: gap of unknown length
22551 24071: contig of 1820 bp in length
24071 25140: gap of unknown length
25141 25220: contig of 990 bp in length
25221 26330: gap of unknown length
26331 26410: contig of 1110 bp in length
26410: gap of unknown length
```

```
TITLE
JOURNAL
REFERENCE
AUTHORS
Richardson,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
2 (bases 1 to 74523)
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richardson,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Direct Submission
Submitted (23-NOV-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 69 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
TITLE
JOURNAL
COMMENT
1317: contig of 1317 bp in length
1318 1397: gap of unknown length
1398 1995: contig of 598 bp in length
1996 2075: gap of unknown length
2076 2858: contig of 783 bp in length
2859 2938: gap of unknown length
2939 3636: contig of 698 bp in length
3637 3716: gap of unknown length
3717 5148: contig of 1431 bp in length
5148 5228: gap of unknown length
5228 5988: contig of 760 bp in length
5988 6067: gap of unknown length
6068 6769: contig of 702 bp in length
6770 6849: gap of unknown length
6850 7674: contig of 825 bp in length
7675 7754: gap of unknown length
7755 8519: contig of 765 bp in length
8520 8599: gap of unknown length
8600 9488: contig of 889 bp in length
9489 9569: gap of unknown length
9569 10567: contig of 999 bp in length
10568 10648: gap of unknown length
10648 12209: contig of 1562 bp in length
12210 12289: gap of unknown length
12290 13010: contig of 721 bp in length
13011 13090: gap of unknown length
13091 14499: contig of 1409 bp in length
14500 14580: gap of unknown length
14580 16194: contig of 1614 bp in length
16194 16274: gap of unknown length
16274 17532: contig of 1259 bp in length
17533 17613: gap of unknown length
17613 19098: contig of 1486 bp in length
19099 19178: gap of unknown length
19179 20704: contig of 1526 bp in length
20705 20785: gap of unknown length
20785 22171: contig of 1386 bp in length
22171 22551: gap of unknown length
22551 24071: contig of 1820 bp in length
24071 25140: gap of unknown length
25141 25220: contig of 990 bp in length
25221 26330: gap of unknown length
26331 26410: contig of 1110 bp in length
26410: gap of unknown length
```

```
* * 26411 27441: contig of 1031 bp in length
* * 27442 27521: gap of unknown length
* * 27522 28939: contig of 1418 bp in length
* * 28940 29019: gap of unknown length
* * 29020 29995: contig of 976 bp in length
* * 29996 30075: gap of unknown length
* * 30076 31915: contig of 1840 bp in length
* * 31916 31995: gap of unknown length
* * 31996 33705: contig of 1710 bp in length
* * 33706 33785: gap of unknown length
* * 33786 35570: contig of 1785 bp in length
* * 35571 35650: gap of unknown length
* * 35651 37849: contig of 2199 bp in length
* * 37850 37929: gap of unknown length
* * 37930 39907: contig of 1978 bp in length
* * 39908 39987: gap of unknown length
* * 39988 42431: contig of 2444 bp in length
* * 42432 42511: gap of unknown length
* * 42512 44943: contig of 2432 bp in length
* * 44944 45023: gap of unknown length
* * 45024 48697: contig of 3674 bp in length
* * 48698 48777: gap of unknown length
* * 48778 49333: contig of 556 bp in length
* * 49334 49413: gap of unknown length
* * 49414 50232: contig of 818 bp in length
* * 50233 50311: gap of unknown length
* * 50312 50850: contig of 539 bp in length
* * 50851 50930: gap of unknown length
* * 50931 51690: contig of 680 bp in length
* * 51611 51690: gap of unknown length
* * 51691 52318: contig of 548 bp in length
* * 52319 52399: gap of unknown length
* * 52400 52952: contig of 634 bp in length
* * 52953 53032: gap of unknown length
* * 53033 53624: contig of 592 bp in length
* * 53625 53704: gap of unknown length
* * 53705 54356: contig of 652 bp in length
* * 54357 54436: gap of unknown length
* * 54437 55076: contig of 640 bp in length
* * 55077 55156: gap of unknown length
* * 55157 55788: contig of 632 bp in length
* * 55789 55868: gap of unknown length
* * 55869 56673: contig of 805 bp in length
* * 56674 56753: gap of unknown length
* * 56754 57401: contig of 648 bp in length
* * 57402 57481: gap of unknown length
* * 57482 58088: contig of 607 bp in length
* * 58089 58168: gap of unknown length
* * 58169 58624: contig of 456 bp in length
* * 58625 58704: gap of unknown length
* * 58705 59325: contig of 621 bp in length
* * 59326 59405: gap of unknown length
* * 59406 60056: contig of 651 bp in length
* * 60057 60136: gap of unknown length
* * 60137 60826: contig of 690 bp in length
* * 60827 60906: gap of unknown length
* * 60907 61484: contig of 578 bp in length
* * 61485 61564: gap of unknown length
* * 61565 62192: contig of 628 bp in length
* * 62193 62272: gap of unknown length
* * 62273 63066: contig of 794 bp in length
* * 63067 63146: gap of unknown length
* * 63147 63770: contig of 624 bp in length
* * 63771 63850: gap of unknown length
* * 63851 64511: contig of 661 bp in length
* * 64512 64591: gap of unknown length
* * 64592 65201: contig of 610 bp in length
* * 65202 65281: gap of unknown length
* * 65282 65925: contig of 644 bp in length
* * 65926 66005: gap of unknown length
* * 66006 66573: contig of 368 bp in length
* * 66574 66653: gap of unknown length
* * 66654 67411: contig of 758 bp in length
```

```
* * 67412 67491: gap of unknown length
* * 67492 68070: contig of 579 bp in length
* * 68071 68150: gap of unknown length
* * 68151 68865: contig of 715 bp in length
* * 68866 68945: gap of unknown length
* * 68946 69645: contig of 700 bp in length
* * 69646 69725: gap of unknown length
* * 69726 70421: contig of 696 bp in length
* * 70422 70501: gap of unknown length
* * 70502 70949: contig of 448 bp in length
* * 70950 71029: gap of unknown length
* * 71030 71646: contig of 617 bp in length
* * 71647 71726: gap of unknown length
* * 71727 72365: contig of 639 bp in length
* * 72366 72445: gap of unknown length
* * 72446 73207: contig of 762 bp in length
* * 73208 73287: gap of unknown length
* * 73288 73826: contig of 539 bp in length
* * 73827 73906: gap of unknown length
* * 73907 74523: contig of 617 bp in length.
```

FEATURES  
source

```
1. .74523
   /organism="Drosophila melanogaster"
   /strain="y; cn bw sp"
   /db_xref="taxon:7227"
   /chromosome="3"
   /map="92D-92D"
   /clone="BACR38H07 (D1203) RPCI-98 38.H.7"
   /clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBACE3.6"
```

BASE COUNT 19151 a 15273 c 15080 g 19573 t 5446 others  
ORIGIN

Query Match 1-9%; Score 22; DB 44; Length 74523;

Best Local Similarity 100.0%; Pred. No. 0.63; Length 74523;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 cgctgctgctgcgcgcgcacc 62

Db 47179 CGCTGCTGCTGCGCGCCACC 47158

## RESULT 8

AC010199

LOCUS

DEFINITION Homo sapiens clone RP11-587P7, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 70  
unordered pieces.

ACCESSION AC010199

VERSION AC010199.3 GI:6513886

KEYWORDS HTG; HTGS\_PHASE1.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 109725)  
Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,  
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,  
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,  
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,  
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,  
Forcum-Tansey,J., Frantz,P., Ganes,R., Gorrell,J.H., Gorrell,L.L.,  
Guevara-W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,  
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,  
Kelly,S., Kondajewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,  
Lichter,O., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,  
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,  
Morris,S., Nash,S., Neilson,A., Nguyen,R., Nguyen,N., Nguyen,S.,  
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,  
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,  
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,  
Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,O., Wabba,M.,

Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A.,  
 Worley, K., Wren, J., Wrenford, G., Yu, W., Zhou, X., Nelson, D. and  
 Gibbs, R.  
 Direct Submission  
 2 (bases 1 to 109725)  
 Worley, K.C.  
 Direct Submission  
 Submitted (15-SEP-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Dec 3, 1999 this sequence version replaced gi:6087857.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Summary Statistics  
 Sequencing vector: M13; L08821  
 Chemistry: Dye-primer Bodipy: 14% of reads  
 Chemistry: Dye-terminator Big Dye: 84% of reads  
 Assembly program: Phrap; version 0.980611  
 Consensus quality: 190820 bases at least Q40  
 Consensus quality: 203698 bases at least Q30  
 Consensus quality: 210538 bases at least Q20  
 Estimated insert size: 193000; agarose-fp estimation  
 Estimated insert size: 210705; sum-of-contigs estimation  
 Quality coverage: 3.3x in Q20 bases; agarose-fp estimation  
 Quality coverage: 3.1x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 70 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 4276: contig of 4276 bp in length  
 \* 4277 4296: gap of unknown length  
 \* 4297 7908: contig of 3612 bp in length  
 \* 7909 7929: gap of unknown length  
 \* 7929 11528: contig of 3600 bp in length  
 \* 11529 11548: gap of unknown length  
 \* 11549 15039: contig of 3491 bp in length  
 \* 15040 15059: gap of unknown length  
 \* 15060 18193: contig of 3134 bp in length  
 \* 18194 18213: gap of unknown length  
 \* 18214 21316: contig of 3103 bp in length  
 \* 21317 21336: gap of unknown length  
 \* 21337 24270: contig of 2934 bp in length  
 \* 24271 24290: gap of unknown length  
 \* 24291 27202: contig of 2912 bp in length  
 \* 27203 27222: gap of unknown length  
 \* 27223 30045: contig of 2823 bp in length  
 \* 30046 30065: gap of unknown length  
 \* 30066 32710: contig of 2645 bp in length  
 \* 32711 32730: gap of unknown length  
 \* 32731 35211: contig of 2481 bp in length  
 \* 35212 35231: gap of unknown length  
 \* 35232 37697: contig of 2466 bp in length  
 \* 37698 37717: gap of unknown length  
 \* 37718 40007: contig of 2290 bp in length  
 \* 40008 40027: gap of unknown length  
 \* 40028 42272: contig of 2245 bp in length  
 \* 42273 42292: gap of unknown length  
 \* 42293 44399: contig of 2107 bp in length  
 \* 44400 44419: gap of unknown length  
 \* 44420 46460: contig of 2041 bp in length  
 \* 46461 46480: gap of unknown length  
 \* 46481 48406: contig of 1926 bp in length  
 \* 48407 48426: gap of unknown length  
 \* 48427 50350: contig of 1924 bp in length

50351 50370: gap of unknown length  
 50371 52202: contig of 1832 bp in length  
 52203 52222: gap of unknown length  
 52223 54046: contig of 1824 bp in length  
 54047 54066: gap of unknown length  
 54067 55863: contig of 1797 bp in length  
 55864 55883: gap of unknown length  
 55884 57533: contig of 1670 bp in length  
 57534 57573: gap of unknown length  
 57574 59214: contig of 1641 bp in length  
 59215 59234: gap of unknown length  
 59235 60791: contig of 1557 bp in length  
 60792 60811: gap of unknown length  
 60812 62362: contig of 1551 bp in length  
 62363 63886: gap of unknown length  
 63887 63906: gap of unknown length  
 63907 65395: contig of 1490 bp in length  
 65397 65416: gap of unknown length  
 65417 66896: contig of 1480 bp in length  
 66897 68395: gap of unknown length  
 68396 68415: gap of unknown length  
 68416 69845: contig of 1431 bp in length  
 69847 69866: gap of unknown length  
 69867 71240: contig of 1374 bp in length  
 71241 72628: gap of unknown length  
 72629 72648: gap of unknown length  
 72649 73988: contig of 1340 bp in length  
 73989 74008: gap of unknown length  
 74009 75331: contig of 1323 bp in length  
 75332 75351: gap of unknown length  
 75352 76662: contig of 1311 bp in length  
 76663 76682: gap of unknown length  
 76683 77972: contig of 1290 bp in length  
 77973 77992: gap of unknown length  
 77993 79275: contig of 1283 bp in length  
 79276 79295: gap of unknown length  
 79296 80546: contig of 1250 bp in length  
 80547 80565: gap of unknown length  
 80566 81777: contig of 1212 bp in length  
 81778 82987: gap of unknown length  
 82988 83007: contig of 1190 bp in length  
 83008 84067: gap of unknown length  
 84068 85146: gap of unknown length  
 85147 85166: gap of unknown length  
 85167 86208: contig of 1040 bp in length  
 86209 86226: gap of unknown length  
 86227 87250: contig of 1024 bp in length  
 87251 88215: gap of unknown length  
 88216 88235: gap of unknown length  
 88236 89168: contig of 933 bp in length  
 89169 90095: gap of unknown length  
 90096 90115: gap of unknown length  
 90116 91021: contig of 906 bp in length  
 91022 91041: gap of unknown length  
 91042 91944: contig of 903 bp in length  
 91945 92855: gap of unknown length  
 92856 92875: gap of unknown length  
 92876 93763: contig of 888 bp in length  
 93764 94680: gap of unknown length  
 94681 95574: contig of 877 bp in length  
 95575 95574: gap of unknown length  
 95575 96447: contig of 874 bp in length  
 96448 96467: gap of unknown length  
 96469 96480: contig of 873 bp in length  
 96481 96499: gap of unknown length

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT





```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
10523: contig of 10523 bp in length
gap of unknown length
10524 13538: contig of 3015 bp in length
gap of unknown length
13539 17077: contig of 3539 bp in length
gap of unknown length
17078 20904: contig of 3827 bp in length
gap of unknown length
20905 25000: contig of 4096 bp in length
gap of unknown length
25001 30177: contig of 5177 bp in length
gap of unknown length
30178 36463: contig of 6286 bp in length
gap of unknown length
36464 43571: contig of 7108 bp in length
gap of unknown length
43572 53182: contig of 9611 bp in length
gap of unknown length
53183 62954: contig of 9772 bp in length
gap of unknown length
62955 72140: contig of 9186 bp in length
gap of unknown length
72141 82556: contig of 10416 bp in length
gap of unknown length
82557 96547: contig of 13991 bp in length
gap of unknown length
96548 109713: contig of 13166 bp in length
gap of unknown length
109714 125535: contig of 15822 bp in length
gap of unknown length
125536 127207: contig of 1672 bp in length.
Location/Qualifiers
1. 127207
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="44_K_6"
/clone.lib="RPCI-11 Human Male BAC"
37605 a 26430 c 25609 g 36876 t 687 others

Query Match 1.9%; Score 22; DB 33; Length 127207;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1152 cttccatctcctcaaaaaaaaaa 1173
Db 20903 CTTCCATCTCAAAAAAAAAA 20924

RESULT 11
AC010193
LOCUS Homo sapiens clone RP11-469G7, *** SEQUENCING IN PROGRESS ***, 42
DEFINITION Homo sapiens clone RP11-469G7, *** SEQUENCING IN PROGRESS ***, 42
ACCESSION AC010193
VERSION AC010193.5 GI:6513887
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 132398)
Muzny,D.M., Adams,C., Bailey,M., Barbarta,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,

```

```

Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugang,R.,
Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.
Direct Submission
Unpublished
REFERENCE 2 (bases 1 to 132398)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 3, 1999 this sequence version replaced gi:6087861.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 95% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 145481 bases at least Q40
Consensus quality: 153270 bases at least Q30
Estimated insert size: 141000; agarose-fp estimation
Estimated insert size: 156282; sum-of-contigs estimation
Quality coverage: 4.4x in Q20 bases; agarose-fp estimation
Quality coverage: 4.0x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 42 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 9500: contig of 9500 bp in length
* 9520: gap of unknown length
* 9521 18550: contig of 9030 bp in length
* 18551 18570: gap of unknown length
* 18571 27309: contig of 8739 bp in length
* 27310 27329: gap of unknown length
* 27330 35858: contig of 8529 bp in length
* 35859 35878: gap of unknown length
* 35879 43932: contig of 8054 bp in length
* 43933 43952: gap of unknown length
* 43953 51800: contig of 7848 bp in length
* 51801 51820: gap of unknown length
* 51821 59385: contig of 7565 bp in length
* 59386 59405: gap of unknown length
* 59406 64832: contig of 5427 bp in length
* 64833 64852: gap of unknown length
* 64853 69885: contig of 5034 bp in length
* 69887 69907: gap of unknown length
* 69908 74250: contig of 4344 bp in length
* 74251 74270: gap of unknown length
* 74271 78045: contig of 3775 bp in length
* 78046 78065: gap of unknown length
* 78066 81374: contig of 3289 bp in length
* 81375 81376: gap of unknown length
* 81377 84292: contig of 2918 bp in length
* 84293 84312: gap of unknown length

```



```
* 84313 87107: contig of 2795 bp in length
* 87108 87127: gap of unknown length
* 87128 89834: contig of 2707 bp in length
* 89835 89854: gap of unknown length
* 89855 92537: contig of 2683 bp in length
* 92538 92557: gap of unknown length
* 92558 95081: contig of 2524 bp in length
* 95082 95101: gap of unknown length
* 95102 97602: contig of 2501 bp in length
* 97603 97623: gap of unknown length
* 97623 100026: contig of 2404 bp in length
* 100027 100046: gap of unknown length
* 100047 102144: contig of 2098 bp in length
* 102145 102164: gap of unknown length
* 102165 104173: contig of 2011 bp in length
* 104176 104195: gap of unknown length
* 104196 106187: contig of 1992 bp in length
* 106188 106207: gap of unknown length
* 106208 108061: contig of 1854 bp in length
* 108062 108081: gap of unknown length
* 108082 109861: contig of 1780 bp in length
* 109862 109881: gap of unknown length
* 109882 111483: contig of 1602 bp in length
* 111484 111503: gap of unknown length
* 111504 113104: contig of 1601 bp in length
* 113105 113124: gap of unknown length
* 113125 114707: contig of 1583 bp in length
* 114708 114727: gap of unknown length
* 114728 116274: contig of 1547 bp in length
* 116275 116294: gap of unknown length
* 116295 117823: contig of 1528 bp in length
* 117824 117843: gap of unknown length
* 117844 119267: contig of 1425 bp in length
* 119268 119287: gap of unknown length
* 119288 120650: contig of 1363 bp in length
* 120651 120671: gap of unknown length
* 120671 122031: contig of 1361 bp in length
* 122032 122051: gap of unknown length
* 122052 123331: contig of 1280 bp in length
* 123332 123351: gap of unknown length
* 123352 124571: contig of 1220 bp in length
* 124572 124591: gap of unknown length
* 124592 125777: contig of 1186 bp in length
* 125778 125797: gap of unknown length
* 125798 126977: contig of 1182 bp in length
* 126978 126998: gap of unknown length
* 126999 127000: contig of 999 bp in length
* 127000 127999: gap of unknown length
* 127999 128019: gap of unknown length
* 128019 128963: contig of 945 bp in length
* 128964 128983: gap of unknown length
* 128984 129851: contig of 868 bp in length
* 129852 129871: gap of unknown length
* 129872 130715: contig of 844 bp in length
* 130716 130735: gap of unknown length
* 130736 131563: contig of 828 bp in length
* 131564 131583: gap of unknown length
* 131584 132398: contig of 815 bp in length.
FEATURES
    source
    1..132398
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="RP11-469G7"
BASE COUNT 34716 a 30704 c 31220 g 34876 t 882 others
ORIGIN
    1..132398
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="RP11-469G7"
```

```
Query Match 1.9% Score 22; DB 44; Length 132398;
Best Local Similarity 100.0%; Pred. No. 0.57;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1152 ctccatctcaaaaaa 1173
|||||
Db 113869 CTTCATCTCAAAAAA 113890
```

## RESULT 12

HUAC003049

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

HUAC003049 134450 bp DNA PRI 20-MAR-1998  
Homo sapiens Chromosome 16 BAC clone CIT987SK-A-485G10, complete  
sequence.  
AC003049  
AC003049.1 GI:2979538  
HTG.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 134450)  
Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,  
Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.  
Homo sapiens Chromosome 16 BAC clone CIT987SK-A-485G10  
Unpublished  
2 (bases 1 to 134450)  
Adams,M.D. and Loftus,B.J.  
Direct Submission  
Submitted (03-NOV-1997) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA, Email:  
bjloftus@tigr.org  
3 (bases 1 to 134450)  
Adams,M.D. and Loftus,B.J.  
Direct Submission  
Submitted (20-MAR-1998) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
On Mar 20, 1998 this sequence version replaced gi:2909695.  
Address all correspondence to: Mark Adams The Institute for Genomic  
Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail  
address: mdadams@tigr.org. The orientation of the sequence is  
from SP6 end to T7 end. Genes were identified by a combination of  
five methods including: XGRAIL (available by anonymous ftp from  
arthur.epm.ornl.gov), Genefinder (Phil Green, University of  
Washington), Genscan (Chris Burge,  
http://gnomic.stanford.edu/~chris/GENSCANW.html) searches of the  
complete sequence against a peptide database, and the Human gene  
Index database at TIGR (http://www.tigr.org/tdb/hgi.html).  
Genes without peptide homology having spliced EST hits are termed  
'unknown gene product'. Genes encoding tRNAs are predicted by  
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).

## FEATURES

source

1..134450

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="16"

/map="16p11.2-p12"

/clone="A-485G10"

43645..43695

/note="50 N's represent a physical gap containing an ATC

repeat of approximately 800bp"

BASE COUNT 40427 a 28300 c 29112 g 36561 t 50 others

ORIGIN

Query Match 1.9% Score 22; DB 11; Length 134450;  
Best Local Similarity 100.0%; Pred. No. 0.57;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1152 ctccatctcaaaaaa 1173
|||||
Db 30834 CTTCATCTCAAAAAA 30855
```

## RESULT 13

AC004477

LOCUS

DEFINITION

ACCESSION

VERSION

AC004477 135505 bp DNA PRI 29-OCT-1998  
Homo sapiens chromosome 17, clone HRPC890E16, complete sequence.  
AC004477  
AC004477.1 GI:3688107

```

KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS     Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       1 (bases 1 to 135505)
JOURNAL     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
REFERENCE   Homo sapiens chromosome 17, clone HRPC890E16
AUTHORS     2 (bases 1 to 135505)
REFERENCE   Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P.,
AUTHORS     Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R.,
Benn,J., Boatn,C., Boutwell,C., Brown,A., Byrne,S., Cantu,C.,
Castle,A., Cerny,J., Cooke,P., Daly,M.J., Depayre,E., Devon,K.,
Dewar,K., Donelan,L., DuRette,B., Etemadi,S., Ferreira,P.,
FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Gray,D.,
Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L.,
Kann,L., Linton,L., MacDonald,P., Marquis,N., McEwan,P., McGurk,A.,
Meidrim,J., Molla,M., Morris,W., Morrow,J., Nachman,A., Nahf,R.,
Naylor,J., O'Connor,T., Pavlin,B., Peterson,K., Riley,R.,
Roberts,D., Rollins,G., Rossello,R., Roy,A., Shyam,R.,
Stange-Thomann,N., Stilwell,J., Stone,C., Strickland,C.,
Subramanian,A., Sydney,K., Tang,L., Vassiliev,H., Vo,A., Wagner,A.,
Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
TITLE       Submitted (26-MAR-1998) Whitehead Institute/MIT Center for Genome
JOURNAL     Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 135505)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Collins,D., Depayre,E., Devon,K.,
Collimore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heaford,A.,
Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lehoczyk,J., MacDonald,P., Marquis,N., McEwan,P.,
McGurk,A., McKernan,K., Meidrim,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Testaye,S., Tichovolsky,N., Torruella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
TITLE       Submitted (02-OCT-1998) Whitehead Institute/MIT Center for Genome
JOURNAL     Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4 (bases 1 to 135505)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Collins,D., Depayre,E., Devon,K.,
Collimore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heaford,A.,
Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lehoczyk,J., MacDonald,P., Marquis,N., McEwan,P.,
McGurk,A., McKernan,K., Meidrim,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Testaye,S., Tichovolsky,N., Torruella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
TITLE       Submitted (29-OCT-1998) Whitehead Institute/MIT Center for Genome
JOURNAL     Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 2, 1998 this sequence version replaced gi:3687291.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES   Location/Qualifiers
SOURCE     1..135505
           /organism="Homo sapiens"
db_xref="taxon:9606"
/clone="HRPC890E16"
/clone_lib="Peter de Jong/ human PAC library"
/map="17"
/chromosome="17"
293..611
/rpt_family="AluSg"
/rpt_family="AluSg"
4777..5071
/rpt_family="AluSp"
complement(5641..5739)
/rpt_family="MIR"
complement(5771..6242)
/rpt_family="MLTID"
6431..6617
/rpt_family="LTR16A"
complement(6716..7004)
/rpt_family="AluSx"
7051..7381
/rpt_family="MLTIC"
7051..7381
/rpt_family="MLTIC"
complement(7382..7682)
/rpt_family="AluY"
7683..7757
/rpt_family="MLTIC"
complement(7824..7867)
/rpt_family="Polypurine"
complement(7870..8002)
/rpt_family="FLAM_C"
complement(8127..8255)
/rpt_family="FRAM"
8257..8305
/rpt_family="U17"
8318..8611
/rpt_family="L2"
complement(8723..9005)
/rpt_family="AluSx"
complement(9737..9889)
/rpt_family="L2"
9890..10197
/rpt_family="AluSg"
complement(10198..10214)
/rpt_family="L2"
10541..10724
/rpt_family="MIR"
complement(10725..11018)
/rpt_family="AluSx"
complement(11044..11354)
/rpt_family="AluY"
11357..11503
/rpt_family="MIR"
complement(12451..12627)
/rpt_family="AluSx"
complement(12628..12936)
/rpt_family="AluSg"
complement(12937..13072)
/rpt_family="AluSx"
complement(13162..13267)
/rpt_family="MIR"
13397..13708
/rpt_family="AluY"
13736..14395
/rpt_family="L2"
complement(14441..14526)
/rpt_family="MER45"
15808..16131
/rpt_family="AluSp"
complement(16136..16210)
/rpt_family="MIR"
complement(16773..16891)
/rpt_family="L2"
complement(16837..17036)
/rpt_family="MIR"
complement(18992..19200)
/rpt_family="MIR"

```

```

repeat_region complement(20490..20784)
/rpt_family="AluJb"
repeat_region complement(21097..21261)
/rpt_family="AluY"
repeat_region complement(21262..21568)
/rpt_family="AluYab"
repeat_region complement(21569..21699)
/rpt_family="AluY"
repeat_region complement(22191..22492)
/rpt_family="AluSp"
repeat_region complement(23193..23776)
/rpt_family="L2"
repeat_region complement(24385..24689)
/rpt_family="AluJo"
repeat_region 27706..27809
/rpt_family="MIR"
repeat_region complement(28745..29001)
/rpt_family="AluSq"
repeat_region complement(29007..29309)
/rpt_family="AluSq"
repeat_region complement(29319..29447)
/rpt_family="FLAM_C"
repeat_region 29541..29730
/rpt_family="MIR"
repeat_region 29731..30036
/rpt_family="AluSq"
repeat_region 30037..30072
/rpt_family="MIR"
repeat_region 30128..30411
/rpt_family="AluJo"
repeat_region 31164..31257
/rpt_family="AluJ/FLAM"
repeat_region 32320..32618
/rpt_family="AluSq"
repeat_region 32629..32761
/rpt_family="AluSq/x"
repeat_region complement(33193..33294)
/rpt_family="LMB3"
repeat_region complement(34367..34665)
/rpt_family="AluX"
repeat_region complement(35444..35536)
/rpt_family="purine-rich"
repeat_region 36104..36411
/rpt_family="AluSq"
repeat_region 36480..36629
/rpt_family="MIR"
repeat_region 36832..36915
/rpt_family="MIR"
repeat_region complement(36916..36946)
/rpt_family="(TAA)n"

```

```

Query Match 1.9%; Score 22; DB 11; Length 135505;
Best Local Similarity 100.0%; Pred. No. 0.57;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1152 cttccatctcaaaaaaaaaa 1173
|||||
Db 107202 CTCCATCTCAAAAAAAAAA 107223

```

```

RESULT 14
AC011163/c AC011163 139149 bp DNA HTG 01-DEC-1999
LOCUS Homo sapiens clone RP11-10A10, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC011163
ACCESSION AC011163
VERSION AC011163.3 GI:6492521
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 139149)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

```

# TITLE JOURNAL REFERENCE AUTHORS

Homo sapiens, clone RP11-10A10  
Unpublished  
2 (bases 1 to 139149)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,  
Balwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Boukhgalter, B.,  
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,  
Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,  
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,  
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
Lewoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,  
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Melarim, J.,  
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,  
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Dec 1, 1999 this sequence version replaced gi:6425748.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submission@genome.wi.mit.edu  
----- Project Information  
Center project name: L3028  
Center clone name: 10\_A\_10  
-----

# TITLE JOURNAL COMMENT

\* NOTE: This record contains 163 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved. 794: contig of 794 bp in length  
\* I gap of unknown length  
\* 795 1593: contig of 799 bp in length  
\* gap of unknown length  
\* 1594 2362: contig of 769 bp in length  
\* gap of unknown length  
\* 2363 3187: contig of 825 bp in length  
\* gap of unknown length  
\* 3188 3981: contig of 794 bp in length  
\* gap of unknown length  
\* 3982 4793: contig of 812 bp in length  
\* gap of unknown length  
\* 4794 5626: contig of 833 bp in length  
\* gap of unknown length  
\* 5627 6446: contig of 820 bp in length  
\* gap of unknown length  
\* 6447 7239: contig of 793 bp in length  
\* gap of unknown length  
\* 7240 8070: contig of 831 bp in length  
\* gap of unknown length  
\* 8071 8902: contig of 832 bp in length  
\* gap of unknown length  
\* 8903 9715: contig of 813 bp in length  
\* gap of unknown length  
\* 9716 10522: contig of 807 bp in length  
\* gap of unknown length  
\* 10523 11339: contig of 817 bp in length  
\* gap of unknown length  
\* 11340 12156: contig of 817 bp in length

```
* * 12157 gap of unknown length
* * 12972: contig of 816 bp in length
* * gap of unknown length
* * 12973: contig of 830 bp in length
* * gap of unknown length
* * 13803: contig of 809 bp in length
* * gap of unknown length
* * 14612: contig of 833 bp in length
* * gap of unknown length
* * 15445: contig of 815 bp in length
* * gap of unknown length
* * 16260: contig of 829 bp in length
* * gap of unknown length
* * 17089: contig of 819 bp in length
* * gap of unknown length
* * 17908: contig of 812 bp in length
* * gap of unknown length
* * 18720: contig of 822 bp in length
* * gap of unknown length
* * 19542: contig of 812 bp in length
* * gap of unknown length
* * 20354: contig of 825 bp in length
* * gap of unknown length
* * 21179: contig of 817 bp in length
* * gap of unknown length
* * 21996: contig of 812 bp in length
* * gap of unknown length
* * 22808: contig of 823 bp in length
* * gap of unknown length
* * 23631: contig of 818 bp in length
* * gap of unknown length
* * 24449: contig of 817 bp in length
* * gap of unknown length
* * 25266: contig of 806 bp in length
* * gap of unknown length
* * 26072: contig of 821 bp in length
* * gap of unknown length
* * 26893: contig of 815 bp in length
* * gap of unknown length
* * 27708: contig of 811 bp in length
* * gap of unknown length
* * 28519: contig of 818 bp in length
* * gap of unknown length
* * 29337: contig of 803 bp in length
* * gap of unknown length
* * 30140: contig of 814 bp in length
* * gap of unknown length
* * 30954: contig of 824 bp in length
* * gap of unknown length
* * 31778: contig of 815 bp in length
* * gap of unknown length
* * 32593: contig of 827 bp in length
* * gap of unknown length
* * 33420: contig of 802 bp in length
* * gap of unknown length
* * 34222: contig of 813 bp in length
* * gap of unknown length
* * 35035: contig of 768 bp in length
* * gap of unknown length
* * 35803: contig of 790 bp in length
* * gap of unknown length
* * 36593: contig of 818 bp in length
* * gap of unknown length
* * 37411: contig of 810 bp in length
* * gap of unknown length
* * 38221: contig of 812 bp in length
* * gap of unknown length
* * 39033: contig of 816 bp in length
* * gap of unknown length
* * 39849: contig of 823 bp in length
* * gap of unknown length
* * 40672: contig of 826 bp in length
* * gap of unknown length
```

```
* * 41498 42307: contig of 810 bp in length
* * gap of unknown length
* * 42308 43127: contig of 820 bp in length
* * gap of unknown length
* * 43128 43943: contig of 816 bp in length
* * gap of unknown length
* * 43944 44771: contig of 828 bp in length
* * gap of unknown length
* * 44772 45599: contig of 828 bp in length
* * gap of unknown length
* * 45600 46399: contig of 800 bp in length
* * gap of unknown length
* * 46400 47229: contig of 830 bp in length
* * gap of unknown length
* * 47230 48034: contig of 805 bp in length
* * gap of unknown length
* * 48035 48858: contig of 824 bp in length
* * gap of unknown length
* * 48859 49665: contig of 807 bp in length
* * gap of unknown length
* * 49666 50472: contig of 807 bp in length
* * gap of unknown length
* * 50473 51299: contig of 827 bp in length
* * gap of unknown length
* * 51300 52116: contig of 817 bp in length
* * gap of unknown length
* * 52117 52944: contig of 828 bp in length
* * gap of unknown length
* * 52945 53766: contig of 822 bp in length
* * gap of unknown length
* * 53767 54579: contig of 813 bp in length
* * gap of unknown length
* * 54580 55392: contig of 813 bp in length
* * gap of unknown length
* * 55393 56204: contig of 812 bp in length
* * gap of unknown length
* * 56205 57023: contig of 819 bp in length
* * gap of unknown length
* * 57024 57824: contig of 801 bp in length
* * gap of unknown length
* * 57825 58646: contig of 822 bp in length
* * gap of unknown length
* * 58647 59467: contig of 821 bp in length
```

Query Match 1.9%; Score 22; DB 44; Length 139149;  
Best Local Similarity 100.0%; Pred. No. 0.57;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1152 ctccatctcaaaaaaaaaa 1173  
|||||  
Db 37533 CTCCATCTCAAAAAAAAAA 37512

## RESULT 15

AC007712 139180 bp DNA HTG 20-SEP-1999  
LOCUS Drosophila melanogaster chromosome 3 clone BACR27K02 (D734) RPCI-98  
DEFINITION 27.K.2 map 92D-92D strain y; cn bw sp, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 96 unordered pieces.  
ACCESSION AC007712  
VERSION AC007712.4 GI:5912626  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 139180)  
AUTHORS Ceiniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blaze),R.G.,  
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,

Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleeb, J.M., Park, S.,  
Preiffer, B., Poon, L., Sequeira, A., Sethi, H., Shif, E.,  
Swirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
Rubin, G.M.  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Sequencing of *Drosophila melanogaster*  
2 (bases 1 to 139180)  
Celinker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,  
Butenhorst, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleeb, J.M., Park, S.,  
Preiffer, B., Poon, L., Sequeira, A., Sethi, H., Shif, E.,  
Swirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
Rubin, G.M.  
TITLE  
JOURNAL  
COMMENT  
Direct Submission  
Submitted (03-JUN-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 84-121, Berkeley, CA 94720, USA  
On Sep 20, 1999 this sequence version replaced g1:5670507.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdpa@fruitfly.berkeley.edu](mailto:bdpa@fruitfly.berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 200 bases.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 96 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1  
927: contig of 927 bp in length  
928  
1008  
1843: contig of 836 bp in length  
1844  
1923: gap of unknown length  
1924  
2751: contig of 828 bp in length  
2752  
2831: gap of unknown length  
2832  
3598: contig of 767 bp in length  
3599  
3678: gap of unknown length  
3679  
4675: contig of 997 bp in length  
4676  
4755: gap of unknown length  
4756  
5467: contig of 712 bp in length  
5468  
5547: gap of unknown length  
5548  
6774: contig of 1227 bp in length  
6775  
6854: gap of unknown length  
6855  
8146: contig of 1292 bp in length  
8147  
8226: gap of unknown length  
8227  
9083: contig of 837 bp in length  
9084  
9163: gap of unknown length  
9164  
9992: contig of 829 bp in length  
9993  
10072: gap of unknown length  
10073  
11016: contig of 944 bp in length  
11017  
11096: gap of unknown length  
11097  
11706: contig of 610 bp in length  
11707  
11786: gap of unknown length  
11787  
12679: contig of 893 bp in length  
12680  
12759: gap of unknown length  
12760  
13877: contig of 1118 bp in length  
13878  
13957: gap of unknown length  
13958  
15089: contig of 1132 bp in length  
15090  
15169: gap of unknown length  
15170  
16140: contig of 971 bp in length  
16141  
16220: gap of unknown length  
16221  
17624: contig of 1404 bp in length  
17625  
17704: gap of unknown length  
17705  
18711: contig of 1007 bp in length  
18712  
18791: gap of unknown length  
18792  
20391: contig of 1600 bp in length  
20392  
20471: gap of unknown length  
20472  
21151: contig of 680 bp in length  
21152  
21231: gap of unknown length

21232  
22061: contig of 830 bp in length  
22062  
22141: gap of unknown length  
22142  
23008: contig of 867 bp in length  
23009  
23088: gap of unknown length  
23089  
24489: contig of 1401 bp in length  
24490  
24569: gap of unknown length  
24570  
25425: contig of 856 bp in length  
25426  
25505: gap of unknown length  
25506  
26823: contig of 1318 bp in length  
26824  
26903: gap of unknown length  
26904  
27914: contig of 1011 bp in length  
27915  
27994: gap of unknown length  
27995  
30546: contig of 2552 bp in length  
30547  
30626: gap of unknown length  
30627  
32014: contig of 1388 bp in length  
32015  
32094: gap of unknown length  
32095  
33282: contig of 1188 bp in length  
33283  
33362: gap of unknown length  
33363  
34203: contig of 840 bp in length  
34204  
34282: gap of unknown length  
34283  
35857: contig of 1575 bp in length  
35858  
35937: gap of unknown length  
35938  
36945: contig of 1008 bp in length  
36946  
37025: gap of unknown length  
37026  
37975: contig of 950 bp in length  
37976  
38055: gap of unknown length  
38056  
39261: contig of 1206 bp in length  
39262  
39341: gap of unknown length  
39342  
40765: contig of 1424 bp in length  
40766  
40845: gap of unknown length  
40846  
42658: contig of 1813 bp in length  
42659  
42738: gap of unknown length  
42739  
43568: contig of 830 bp in length  
43569  
43648: gap of unknown length  
43649  
46246: contig of 2598 bp in length  
46247  
46326: gap of unknown length  
46327  
48375: contig of 2049 bp in length  
48376  
48455: gap of unknown length  
48456  
49530: contig of 1075 bp in length  
49531  
52370: gap of unknown length  
52371  
52450: gap of unknown length  
52451  
54403: contig of 1953 bp in length  
54404  
54483: gap of unknown length  
54484  
55638: contig of 1155 bp in length  
55639  
55718: gap of unknown length  
55719  
58110: contig of 2392 bp in length  
58111  
58190: gap of unknown length  
58191  
60654: contig of 2464 bp in length  
60655  
60734: gap of unknown length  
60735  
63525: contig of 2791 bp in length  
63526  
63605: gap of unknown length  
63606  
65964: gap of unknown length  
65965  
67892: contig of 1848 bp in length  
67893  
67972: gap of unknown length  
67973  
71630: contig of 3658 bp in length  
71631  
71710: gap of unknown length  
71711  
74652: contig of 2942 bp in length  
74653  
74732: gap of unknown length  
74733  
78960: contig of 4228 bp in length  
78961  
79040: gap of unknown length  
79041  
82904: contig of 3864 bp in length  
82905  
82984: gap of unknown length  
82985  
87190: contig of 4206 bp in length  
87191  
87270: gap of unknown length  
87271  
93999: contig of 6729 bp in length  
94000  
94073: gap of unknown length  
94080  
108261: contig of 14182 bp in length  
108341: gap of unknown length  
108342  
108968: contig of 627 bp in length  
108969  
109048: gap of unknown length  
109049  
109832: contig of 784 bp in length

Search completed: May 1, 2000, 16:42:42  
Job time: 20662 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 18:51:40 ; Search time 250.34 Seconds  
(without alignments)  
1172.307 Million cell updates/sec

Title: US-09-215-435-117

Perfect score: 1173  
Sequence: 1 gagctgttatggacacg.....tccatctcaaaaaaaaaa 1173

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 311585 seqs, 125096042 residues

Word size : 0

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146	12.4	182	1 X41222	Human secreted pro
2	21	1.8	225	1 Q59849	Human brain Expre
3	21	1.8	302	1 T31590	Probe (BLUR2) for
4	21	1.8	880	1 Q43675	Sequence of Intron
5	21	1.8	1160	1 V84540	Human secreted pro
6	21	1.8	1664	1 Q33391	Type 1 and 2 seque
7	21	1.8	1664	1 T95069	Genomic region con
8	21	1.8	2116	1 N50499	Pleiotropic regula
9	21	1.8	2708	1 V69220	Human G-protein co
10	21	1.8	4742	1 Q33392	ASM genomic DNA. P
11	21	1.8	4742	1 T95070	Human acid sphingo
12	21	1.8	5931	1 Q43701	Sequence of the hu
13	21	1.8	13865	1 V40401	Human tissue facto
14	21	1.8	16891	1 X37084	MEFV gene sequence
15	21	1.8	235033	1 V57926	Hereditary haemoch
16	21	1.8	237326	1 V57903	Hereditary haemoch
17	20	1.7	95	1 Q75099	Plasmid pOKSC18a c
18	20	1.7	257	1 V90484	EST clone DT802. N
19	20	1.7	280	1 T31594	Probe (BLUR11) for
20	20	1.7	301	1 Q59562	Human brain Expre
21	20	1.7	301	1 Q95197	Simple tandem repe
22	20	1.7	306	1 Q29813	Human brain Expre
23	20	1.7	350	1 V90414	EST clone DM194. N
24	20	1.7	358	1 V68897	DNA molecule encod
25	20	1.7	368	1 V90174	EST clone DF757. N
26	20	1.7	385	1 Q60443	Human brain Expre
27	20	1.7	391	1 V17748	Human breast cance
28	20	1.7	399	1 V88775	EST clone HN72. Ne
29	20	1.7	452	1 V99120	DNA methyltransfer
30	20	1.7	454	1 V44298	Human secreted pro
31	20	1.7	726	1 X03016	Human IL-1ra BAC c
32	20	1.7	743	1 N70812	Sequence encoding
33	20	1.7	773	1 X37506	Human secreted pro

c	34	20	1.7	774	1 T79673	BRCA2 cancer suscep
	35	20	1.7	775	1 X27408	Human secreted pro
	36	20	1.7	779	1 Q63859	API sequence obtd.
	37	20	1.7	788	1 Q01787	Partial sequence o
	38	20	1.7	789	1 X20484	Human secreted pro
	39	20	1.7	807	1 T27152	Human Machado-Jose
	40	20	1.7	811	1 X20485	Human secreted pro
	41	20	1.7	849	1 X27372	Human secreted pro
	42	20	1.7	911	1 X37488	Human secreted pro
	43	20	1.7	990	1 X51748	DNA encoding a hum
	44	20	1.7	998	1 X27362	Human secreted pro
	45	20	1.7	1005	1 X37535	Human secreted pro

#### ALIGNMENTS

##### RESULT 1

X41222  
ID X41222 standard; cDNA; 182 BP.  
AC X41222;  
DT 17-JUN-1999 (first entry)  
DE Human secreted protein 5, EST SEQ ID NO:166.  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
OS Homo sapiens.  
PN WO9906548-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; IB1222.  
PR 01-AUG-1997; US-905135.  
PA (GEST ) GENSET.  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
DR P-PSDB; Y12389.  
PT New nucleic acids encoding human secreted proteins - obtained from  
CDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
kidney, lung, umbilical cord, placenta and colon tissue  
Claim 1: Page 385-386; 824pp; English.  
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human  
secreted proteins, and encode the proteins given in Y12261 to Y12514,  
respectively. The proteins given represent the signal peptide and an  
N-terminal fragment of a secreted protein. The nucleic acid sequences  
can be used for producing secreted human gene products. They can also  
be used to develop products for diagnosis and therapy. The proteins  
obtained may have cytokine activity, cell proliferation/differentiation  
activity, haematopoiesis regulating activity, tissue growth regulating  
activity, reproductive hormone regulating activity, chemotactic/  
chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
ligand activity, anti-inflammatory activity, tumour inhibition activity  
or other activities. The products can be used in forensic, gene therapy  
and chromosome mapping procedures. The sequences can also be used for  
obtaining corresponding promoter sequences. The nucleic acids encoding  
the signal peptide can be used for directing extracellular secretion of  
a polypeptide or the insertion of a polypeptide into a membrane, or  
importing a polypeptide into a cell.  
SQ Sequence 182 BP; 26 A; 78 C; 46 G; 32 T;

Query Match 12.4%; Score 146; DB 1; Length 182;  
Best Local Similarity 100.0%; Pred. No. 4.9e-58;  
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gagctgttatggacacgcttctcgctccagatgcgtccgcgccttcttaacgtgtgtcgcgcgcca 60

Db 36 GAGCTGCTTTATGGACACCGCTTCTCGCGCGGCTCTTAACGCTGTGTGTCGCCGCCCA 95

QY 61 cccctgtataccgcgcacgcgctgctcggtccagatgcgtccgcgccttcttaacgtgtgtcgcgcgcca 120

Db 96 CCCCTGTATACCGCGACCGCATGCTCGTCCAGATGTCGTCGCCGCCCAACGATGC 155





KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 OS Homo sapiens.  
 PN W09854963-A2.  
 PD 10-DEC-1998.  
 PF 04-JUN-1998; U11422.

PR 05-SEP-1997; US-057654.  
 PR 05-SEP-1997; US-057656.  
 PR 05-SEP-1997; US-057760.  
 PR 05-SEP-1997; US-057763.  
 PR 05-SEP-1997; US-057769.  
 PR 05-SEP-1997; US-057774.  
 PR 05-SEP-1997; US-057777.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,  
 PI Fan P, Feng P, Ferrie AM, Fischer CL, Florence C,  
 PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW,  
 PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,  
 PI Shi Y, Sobpet DR, Wei Y, Young P, Yu G, Zeng Z;  
 DR WPI; 99-059865/05.  
 DR P-PSDB; W88663, W88971, W88974, W88975, W88976, W88978.  
 PR New isolated human genes and the secreted polypeptides they encode -  
 PT useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 PS Claim 4; Page 390; 772pp; English.  
 CC The invention relates to nucleic acid sequences (W84411 to W84633)  
 CC encoding human secreted proteins (W88534 to W88756). The secreted protein  
 CC gene sequences are deposited with the ATCC under deposit numbers ATCC  
 CC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,  
 CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host  
 CC cells comprising recombinant vectors containing the nucleic acid  
 CC sequences are used for the recombinant production of the secreted  
 CC proteins. The polynucleotide and amino acid sequences are useful for are  
 CC useful for preventing, treating or ameliorating medical conditions e.g.  
 CC by protein or gene therapy. Pathological conditions can be also diagnosed  
 CC by determining the amount of the new polypeptides in a sample or by  
 CC determining the presence of mutations in the new polynucleotides.  
 CC Specific uses are described for each of the polynucleotides, based on  
 CC which tissues they are most highly expressed in, and include developing  
 CC products for the diagnosis or treatment of cancer, neurodegenerative  
 CC disorders, developmental abnormalities and foetal deficiencies, blood  
 CC disorders, tumours, leukemias, diseases of the immune system, autoimmune  
 CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,  
 CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,  
 CC restenosis, prostate diseases, obesity, disorders involving osteoclasts  
 CC such as osteoporosis, arthritis or malignancies, diseases of testes, lung  
 CC or thymus, digestive/endocrine disorders, infections and AIDS. The  
 CC polypeptides are also useful for identifying their binding partners.  
 CC The present sequence represents a gene encoding a human secreted protein  
 CC (see descriptor line for gene number and clone identification).  
 SQ Sequence 1160 BP; 332 A; 250 C; 313 G; 250 T;

Query Match 1.8%; Score 21; DB 1; Length 1160;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1153 ttccatctcaaaaaa 1173  
 ||||||||||||||||  
 Db 1140 TTCCATCTCAAAAAA 1160

RESULT 6  
 Q33391/C  
 ID Q33391 standard; cDNA; 1664 BP.  
 AC Q33391.  
 DT 06-MAY-1993 (first entry)  
 DE Type 1 and 2 sequence containing region of ASM genomic DNA.  
 KW Acid sphingomyelinase; ASM; type; 1; 2; cDNA library; PCR; primer;  
 KW exon; polymerase chain reaction; amplify; intron; cryptic;  
 KW splice site; consensus sequence; mutation; R496L; deltaR608; L302;  
 KW Neimann-Pick disease; NPD; jewish community; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT primer\_bind 1..19  
 FT /tag= a  
 FT /note= "primer 1"  
 FT primer\_bind 1644..1664  
 FT /tag= b

PR 18-DEC-1997; US-070923.  
 PR 06-JUN-1997; US-048877.  
 PR 06-JUN-1997; US-048881.  
 PR 06-JUN-1997; US-048884.  
 PR 06-JUN-1997; US-048893.  
 PR 06-JUN-1997; US-048896.  
 PR 06-JUN-1997; US-048899.  
 PR 06-JUN-1997; US-048915.  
 PR 06-JUN-1997; US-048949.  
 PR 06-JUN-1997; US-048964.  
 PR 06-JUN-1997; US-048972.  
 PR 06-JUN-1997; US-048974.  
 PR 06-JUN-1997; US-049020.  
 PR 06-JUN-1997; US-049375.  
 PR 05-SEP-1997; US-057628.  
 PR 05-SEP-1997; US-057635.  
 PR 05-SEP-1997; US-057644.  
 PR 05-SEP-1997; US-057647.  
 PR 05-SEP-1997; US-057650.  
 PR 05-SEP-1997; US-057651.  
 PR 05-SEP-1997; US-057657.  
 PR 05-SEP-1997; US-057761.  
 PR 05-SEP-1997; US-057764.  
 PR 05-SEP-1997; US-057770.  
 PR 05-SEP-1997; US-057775.  
 PR 06-JUN-1997; US-048875.  
 PR 06-JUN-1997; US-048878.  
 PR 06-JUN-1997; US-048882.  
 PR 06-JUN-1997; US-048885.  
 PR 06-JUN-1997; US-048889.  
 PR 06-JUN-1997; US-048897.  
 PR 06-JUN-1997; US-048900.  
 PR 06-JUN-1997; US-048916.  
 PR 06-JUN-1997; US-048962.  
 PR 06-JUN-1997; US-048970.  
 PR 06-JUN-1997; US-048974.  
 PR 06-JUN-1997; US-049373.  
 PR 05-SEP-1997; US-057584.  
 PR 05-SEP-1997; US-057629.  
 PR 05-SEP-1997; US-057642.  
 PR 05-SEP-1997; US-057645.  
 PR 05-SEP-1997; US-057648.  
 PR 05-SEP-1997; US-057651.  
 PR 05-SEP-1997; US-057652.  
 PR 05-SEP-1997; US-057657.  
 PR 05-SEP-1997; US-057765.  
 PR 05-SEP-1997; US-057771.  
 PR 05-SEP-1997; US-057776.  
 PR 06-JUN-1997; US-048876.  
 PR 06-JUN-1997; US-048880.  
 PR 06-JUN-1997; US-048882.  
 PR 06-JUN-1997; US-048895.  
 PR 06-JUN-1997; US-048898.  
 PR 06-JUN-1997; US-048901.  
 PR 06-JUN-1997; US-048917.  
 PR 06-JUN-1997; US-048963.  
 PR 06-JUN-1997; US-048971.  
 PR 06-JUN-1997; US-049019.  
 PR 06-JUN-1997; US-049019.  
 PR 05-SEP-1997; US-049374.  
 PR 05-SEP-1997; US-057627.  
 PR 05-SEP-1997; US-057634.  
 PR 05-SEP-1997; US-057643.  
 PR 05-SEP-1997; US-057646.  
 PR 05-SEP-1997; US-057649.



```
FT misc_feature /note= "potential lariat branch point consensus"  
FT 1109..1115  
FT /*tag= p  
FT /note= "potential lariat branch point consensus"  
FT misc_feature 1168..1174  
FT /*tag= q  
FT /note= "potential lariat branch point consensus"  
FT misc_feature 1509..1516  
FT /*tag= r  
FT /note= "potential lariat branch point consensus"  
FT 1573..1579  
FT /*tag= s  
FT /note= "potential lariat branch point consensus"  
FT  
PN US5686240-A.  
PD 11-NOV-1997. 250740.  
PF 27-MAY-1994; US-250740.  
PR 03-MAY-1991; US-695572.  
PA (MOUN ) MOUNT SINAI SCHOOL MEDICINE.  
PI Desnick RJ, Schuchman EH;  
DR WPI: 97-536133/51.  
DE Diagnosing Type A or B Niemann-Pick disease - by detecting recessive  
FT mutation in acid sphingomyelinase gene  
FT Example: Columns 59-62: 58pp; English.  
CC The present sequence is the genomic region containing unique type 1  
CC and 2 specific human acid sphingomyelinase (ASM) regions.  
CC Diagnosing Type A or B Niemann-Pick disease (NPD), or identifying a  
CC person as having the potential to genetically transmit Type A or B  
CC NPD, comprises detecting a recessive mutation in the ASM gene,  
CC which results in an alteration of at least 1 amino acid in the ASM  
CC amino acid sequence. The method is especially useful for prenatal  
CC diagnosis in Ashkenazi Jewish populations. The mutation is  
CC Arg496Leu, deltaArg608, Leu302Pro or fsp330, where fsp330 is a  
CC frame shift mutation comprising a cytosine deletion in ASM codon  
CC 330. The mutations are detected by selectively amplifying mutation  
CC containing portions of the ASM gene by PCR using primers  
CC complementary and identical to a portion of the ASM cDNA sequence,  
CC and sequencing the amplified DNA or subjecting it to a  
CC hybridisation assay using mutation specific probes. The ASM type 1  
CC sequence, or the cDNA sequence encoding it can also be used in the  
CC treatment of NPD.  
SQ Sequence 1664 BP; 329 A; 486 C; 366 G; 483 T;  
  
Query Match 1.8%; Score 21; DB 1; Length 1664;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1153 ttccatctcaaaaaa 1173  
DB 498 TTCCATCTCAAAAAA 478  
  
RESULT 8  
ID N50499 standard; DNA; 2116 BP.  
AC N50499;  
DE 20-JAN-1992 (first entry)  
DE Pleiotropic regulator gene afs B.  
KW Prodigiosin; actinohordine; A-factor.  
OS Streptomyces coelicolor.  
FH Key Location/Qualifiers  
FT cds 532..1740  
FT /*tag= a  
FT /note= "Claim 1"  
FT  
PN J60012985-A.  
PD 23-JAN-1985.  
PF 05-JUL-1983; 122052.  
PR 05-JUL-1983; JP-122052.  
PA (BEPP/) BEPPU T.  
DR WPI: 85-058092/10.  
PT Pleiotropic regulator gene - for distributed-feedback-type laser  
FT for forming diffracting-grid on curved semiconductor layer.
```

```
FT Noabstract Dwg 2b-d/2  
PS Disclosure; Fig 4; 5pp; Japanese.  
CC The gene is responsible for the production of A-factor, a minor  
CC regulator concerned with biosynthesis of streptomycin,  
CC autorisance to the antibiotic and sporulation. The sequence is  
CC also concerned with the regulation of expression of the  
CC antimicrobial substances prodigiosin and actinohordine.  
SQ Sequence 2116 BP; 314 A; 747 C; 780 G; 275 T;  
  
Query Match 1.8%; Score 21; DB 1; Length 2116;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 743 tgggtgaagagcccttgccc 763  
DB 845 TGGTGAAGAGGCCCTTGCCC 825  
  
RESULT 9  
ID V69220/c  
AC V69220;  
DE 27-JAN-1999 (first entry)  
DE Human G-protein coupled receptor polypeptide HCEPT09 encoding cDNA.  
KW G-protein coupled receptor; HCEPT09; gene therapy; immune response;  
KW vaccine; HIV-1; HIV-2; cancer; anorexia; bulimia; asthma; hypertension;  
KW Parkinson's disease; heart failure; hypotension; urinary retention;  
KW osteoporosis; angina pectoris; myocardial infarction; ulcer; allergy;  
KW benign prostatic hypertrophy; psychotic disorder; neurological disorder;  
KW anxiety; schizophrenia; manic depression; delirium; dementia; human;  
KW mental retardation; dyskinesias; Huntington's disease; gene mapping;  
KW Gilles de la Tourette's syndrome; linkage analysis; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT cds 1..1272  
FT /*tag= a  
FT /transl_except= (pos:76..78, aa:Gln)  
FT /transl_except= (pos:82..84, aa:Asn)  
FT /transl_except= (pos:145..147, aa:Tyr)  
FT /transl_except= (pos:271..273, aa:Val)  
FT /product= "HCEPT09 polypeptide"  
FT  
PN EP-877083-A1.  
PD 11-NOV-1998.  
PF 09-APR-1998; 302890.  
PR 05-SEP-1997; US-924174.  
PR 07-MAY-1997; US-045889.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PI Elshourbagy N, Halsey WS, Sathe G, Zhu Y;  
DR WPI: 98-570539/49.  
DR P-PSDB; W81460.  
DE New human G-protein coupled receptor polypeptides and  
FT polynucleotides - useful as diagnostic reagents and for prevention  
FT and treatment of HIV, cancer and Parkinson's disease  
PS Claim 4; Pages 7-9; 24pp; English.  
CC This cDNA encodes a human G-protein coupled receptor (HCEPT09)  
CC polypeptide. The HCEPT09 polypeptide can be recombinantly produced by  
CC transforming a host cell with an expression vector comprising the HCEPT09  
CC nucleic acid. HCEPT09 polypeptides and polynucleotides are useful for  
CC diagnosing susceptibility to diseases by detecting mutations in the  
CC HCEPT09 gene using probes containing the HCEPT09 nucleotide sequence, and  
CC can diagnose diseases associated with HCEPT09 imbalance by determining  
CC HCEPT09 polypeptide or mRNA expression levels. HCEPT09 can be used to  
CC screen for agonists and antagonists which bind the HCEPT09 polypeptide.  
CC These can be used in treatment to activate or inhibit HCEPT09 activity,  
CC in addition to direct administration of antisense sequences to prevent  
CC expression, or HCEPT09-polypeptides to treat conditions associated with a  
CC lack of HCEPT09 protein. Gene therapy may also be used to affect  
CC endogenous HCEPT09 polypeptide production. HCEPT09 antibodies are useful  
CC for inducing an immune response to immunise and prevent diseases, and for  
CC isolating HCEPT09 clones or purifying the polypeptides by affinity  
CC chromatography. HCEPT09 polypeptides can be administered directly or as  
CC a vaccine to inoculate against disease. Diseases diagnosed, prevented or
```

CC treated include HIV-1 and -2; cancers; anorexia; bulimia; asthma;  
CC Parkinson's disease; acute heart failure; hypotension; hypertension;  
CC urinary retention; osteoporosis; angina pectoris; myocardial infarction;  
CC ulcers; allergies; benign prostatic hypertrophy; neurological and  
CC psychotic disorders, including anxiety, schizophrenia, manic depression,  
CC delirium, dementia, severe mental retardation and dyskinesias, such as  
CC Huntington's disease or Gilles de la Tourette's syndrome. The HCEPT09  
CC polypeptide is also useful for mapping the gene to a chromosome, allowing  
CC gene inheritance to be studied through linkage analysis.  
SQ Sequence 2708 BP; 632 A; 770 C; 629 G; 675 T;

Query Match 1.8%; Score 21; DB 1; Length 2708;  
Best Local Similarity 100.0%; Pred. No. 1.3; 0; Indels. 0; Gaps 0;  
Matches 21; Conservative 0; Mismatches 0;  
QY 1153 ttccatctcaaaaaa 1173  
|||||  
Db 2057 TTCCATCTCAAAAAA 2037

RESULT 10  
Q33392/c  
ID Q33392 standard; DNA; 4742 BP.  
AC Q33392;  
DT 06-MAY-1993 (first entry)  
DE ASM genomic DNA.  
KW Acid sphingomyelinase; ASM; type; 1; 2; PCR; primer; amplify; cryptic;  
KW polymerase chain reaction; splice site; mutation; R496L; deltaR608;  
KW L302; Niemann-Pick disease; NPD; Jewish community; ss.  
OS Homo sapiens.

FT Key Location/Qualifiers  
FT exon 124..522  
FT /\*tag= a  
FT /number= Exon 1  
FT 523..987  
FT /\*tag= b  
FT /number= Intron 1  
FT 988..1820  
FT /\*tag= c  
FT /number= Exon 2  
FT 1821..2881  
FT /\*tag= d  
FT /number= Intron 2  
FT 2881..3053  
FT /\*tag= e  
FT /number= Exon 3  
FT /note= "Type 1 sequence"  
FT 3054..3282  
FT /\*tag= f  
FT /number= Intron 3  
FT 3283..3359  
FT /\*tag= g  
FT /number= Exon 4  
FT 3360..3561  
FT /\*tag= h  
FT /number= Intron 4  
FT 3562..3707  
FT /\*tag= i  
FT /number= Exon 5  
FT 3708..3863  
FT /\*tag= j  
FT /number= Intron 5  
FT 3864..4642  
FT /\*tag= k  
FT /number= Exon 6  
FT 55..61  
FT /\*tag= l  
FT /bound\_moety= spl  
FT 211..213  
FT /\*tag= m  
FT /note= "Potential initiation codon"  
FT 307..309

FT /\*tag= n  
FT /note= "Potential initiation codon"  
FT 1821..1860  
FT /\*tag= o  
FT /note= "Type 2"  
FT 1858..1865  
FT /\*tag= p  
FT /function= cryptic donor splice site  
FT 2164..2456  
FT /\*tag= q  
FT /note= "Alu 1 homology region"  
FT 4637..4642  
FT /\*tag= r  
FT complement (693..695)  
FT /\*tag= s  
FT /note= "ORF 1 initiation codon"  
FT complement (1335..1337)  
FT /\*tag= t  
FT /note= "ORF 2 initiation codon"  
FT 2708..2710  
FT /\*tag= u  
FT /note= "ORF 3 initiation codon"  
PN EP-520843-A.  
PD 30-DEC-1992.  
PF 30-APR-1992; 401241.  
PR 03-MAY-1991; US-695472.  
PA (MOUN ) MOUNT SINAI MEDICAL CENT.  
PI Desnick RJ, Schuchman EH;  
WPI: 93-001632/01.  
DR Pure and recombinant acid sphingomyelinase and its nucleic acid -  
PT for treatment and diagnosis of Niemann-Pick disease  
PS Claim 48; Fig 6B; 50pp; English.  
CC This sequence represents the acid sphingomyelinase (ASM) genomic DNA  
CC sequence. This ASM gene sequence was isolated by PCR using primers  
CC constructed from common exonic sequences flanking the type 1 and 2  
CC specific sequences. Certain mutations in the ASM gene ie. R496L,  
CC deltaR608 and L302 have been found to correlate with Niemann-Pick  
CC disease (NPD). See also Q33390-423.  
SQ Sequence 4742 BP; 944 A; 1397 C; 1288 G; 1113 T;

Query Match 1.8%; Score 21; DB 1; Length 4742;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1153 ttccatctcaaaaaa 1173  
|||||  
Db 2189 TTCCATCTCAAAAAA 2169

RESULT 11  
T93070/c  
ID T93070 standard; DNA; 4742 BP.  
AC T93070;  
DT 17-FEB-1998 (first entry)  
DE Human acid sphingomyelinase genomic DNA.  
KW Prenatal diagnosis; Type A; Type B; Niemann-Pick disease;  
KW identification; potential genetic transmitter; detection;  
KW recessive mutation; acid sphingomyelinase; Ashkenazi Jew;  
KW human; treatment; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FH intron 1..123  
FT /\*tag= a  
FT /number= 1  
FT 124..522  
FT /\*tag= b  
FT /number= 1  
FT 523..987  
FT /\*tag= c  
FT /number= 2  
FT 988..1820  
FT /\*tag= d

```

FT /number= 2
FT 1821..2881
FT /tag= e
FT /number= 3
FT 2882..3053
FT /tag= f
FT /number= 3
FT 3054..3282
FT /tag= g
FT /number= 4
FT 3283..3359
FT /tag= h
FT /number= 4
FT 3360..3561
FT /tag= i
FT /number= 5
FT 3562..3707
FT /tag= j
FT /number= 5
FT 3708..3863
FT /tag= k
FT /number= 6
FT 3864..4642
FT /tag= l
FT /number= 6
FT 4643..4742
FT /tag= m
FT /number= 7
FT complement (55..61)
FT /tag= n
FT /note= "Sp1 binding site"
FT 211..213
FT /tag= o
FT /note= "potential initiation codon"
FT 307..309
FT /tag= p
FT /note= "potential initiation codon"
FT complement (693..695)
FT /tag= q
FT /note= "initiation codon for ORF1"
FT 1335..1337
FT /tag= r
FT /note= "initiation codon for ORF2"
FT 1821..1860
FT /tag= s
FT /note= "type 2 specific region"
FT 1858..1865
FT /tag= t
FT /note= "potential donor splice site"
FT complement (2163..2456)
FT /tag= u
FT /note= "Alu 1 homology region"
FT 2708..2710
FT /tag= v
FT /note= "initiation codon for ORF3"
FT
FT US5686240-A.
FT 11-NOV-1997.
FT 27-MAY-1994; US-250740.
FT 27-MAY-1994; US-250740.
FT 03-MAY-1991; US-695572.
FT (MOUN ) MOUNT SINAI SCHOOL MEDICINE.
FT Desnick RJ, Schuchman EH;
FT WPI; 97-556133/51.
FT
FT Diagnosing Type A or B Niemann-Pick disease - by detecting recessive
FT mutation in acid sphingomyelinase gene
FT Example: Columns 61-66; 58pp; English.
FT The present sequence is the human acid sphingomyelinase (ASM)
FT genomic DNA.
FT Diagnosing Type A or B Niemann-Pick disease (NPD), or identifying a
FT person as having the potential to genetically transmit Type A or B
FT NPD, comprises detecting a recessive mutation in the ASM gene,
FT which results in an alteration of at least 1 amino acid in the ASM
FT amino acid sequence. The method is especially useful for prenatal

```

```

CC diagnosis in Ashkenazi Jewish populations. The mutation is
CC Arg496Ieu, Gelta9608, Leu302Pro or fsr330, where fsr330 is a
CC frame shift mutation comprising a cytosine deletion in ASM codon
CC 330. The mutations are detected by selectively amplifying mutation
CC containing portions of the ASM gene by PCR using primers
CC complementary and identical to a portion of the ASM cDNA sequence,
CC and sequencing the amplified DNA or subjecting it to a
CC hybridisation assay using mutation specific probes. The ASM type 1
CC sequence, or the cDNA sequence encoding it can also be used in the
CC treatment of NPD.
CC
SQ Sequence 4742 BP; 944 A; 1397 C; 1288 G; 1113 T;

Query Match 1.8%; Score 21; DB 1; Length 4742;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1153 ttccatctcaaaaaaaaaa 1173
Db 2189 TTCCATCTCAAAAAAAAAA 2169

RESULT 12
Q43701
ID Q43701 standard; DNA; 5931 BP.
AC Q43701;
DT 24-SEP-1993 (first entry)
DE Sequence of the human COL1A1 gene extending from intron 25 to exon
DE 52.
KW Type 1 procollagen; COL1A1; COL2A1; osteoporosis; ss.
OS Homo sapiens.
PN WO9311149-A.
PD 10-JUN-1993.
PR 01-DEC-1992; US-803628.
PR 03-DEC-1991; US-803628.
PA (UYJE-) UNIV JEFFERSON THOMAS.
PI Collige A, Constantinou CD, Pack M, Prockop DJ, Sereda L,
PI Spotila LD, Westerhausen A;
DR WPI; 93-196989/24.
PT Method of determining genetic pre-disposition for osteoporosis -
PT by determining nucleotide sequence of portion of DNA encoding the
PT pro-alpha 1 (I) or pro-alpha 2 (I) gene for type 1 pro-collagen
PT Example; Fig 2; 50pp; English.
PS
CC The example concerns the isolation and characterization of
CC nucleotide sequences of introns 25 to 51 of the Pro-alpha-1 (I) gene.
CC For analysis of sequences extending from intron 25 to exon 40, a
CC genomic fragment of the pro-alpha-1 (I) gene was cloned from
CC cultured skin fibroblasts of a proband with osteogenesis imperfecta.
CC The other sequences were obtained by using primers implicated in
CC Q43663-Q43674 and genomic DNA as a template for the polymerase chain
CC reaction. The data provided the nucleotide sequences that are
CC indicated in Fig. 2 (Q43701). Each of the intron sequences in Fig. 2
CC is separately featured and indexed in Q43675-Q43700.
CC
SQ Sequence 5931 BP; 1147 A; 1814 C; 1625 G; 1344 T;

```

```

Query Match 1.8%; Score 21; DB 1; Length 5931;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1153 ttccatctcaaaaaaaaaa 1173
Db 418 TTCCATCTCAAAAAAAAAA 438

RESULT 13
V40401
ID V40401 standard; DNA; 13865 BP.
AC V40401;
DT 16-OCT-1998 (first entry)
DE Human tissue factor full length genomic DNA sequence.
KW Human; truncated; tissue factor; tTF; TF; tumour; coagulation;
KW blood vessel; Factor VIIa; FVIIa; benign growth; vascularised;

```

KW benign prostatic hypertrophy; malignant; necrosis; angiogenesis;  
KW diabetic retinopathy; restenosis; neovascular glaucoma; psoriasis;  
KW rheumatoid arthritis; ss.  
OS Homo sapiens.  
PN WO9831394-A2.  
PD 23-JUL-1998. U01012.  
PR 20-JAN-1998; US-042427.  
PR 27-MAR-1997; US-035920.  
PR 22-JAN-1997; US-036205.  
PA (TEXA ) UNIV TEXAS SYSTEM.  
PI Gao B, King SW, Thorpe PE;  
DR WPI: 98-413821/35.  
DR P-PSDB: W69613.  
PT Composition containing coagulation-defective tissue factor for  
PT treating, e.g. tumours - useful for, e.g. promoting coagulation in  
PT pro-thrombotic and tumour-associated vasculature, used with, e.g.  
PT factor 7 or anti-cancer agent  
PS Disclosure; Page 185-193; 225pp; English.  
CC A composition has been developed which comprises at least 1 coagulation-  
CC deficient tissue factor (TF) compound that is modified to increase its  
CC biological half-life, but excluding modification that involves attachment  
CC to an antibody (or its antigen-binding region) that binds to a component  
CC (cells, vasculature or stroma) of tumours. Also described in the present  
CC invention are compositions containing any coagulation-deficient TF for  
CC promoting coagulation. The coagulation-deficient TFs are used to promote  
CC coagulation preferentially in prothrombotic vessels, particularly those  
CC associated with: (i) benign growths (e.g. benign prostatic hypertrophy);  
CC (ii) vascularised, malignant tumours of medium or large size (where they  
CC also induce tumour necrosis), or (iii) other disorders that involve  
CC angiogenesis, e.g. diabetic retinopathy, restenosis, neovascular  
CC glaucoma, psoriasis and rheumatoid arthritis. The composition can be  
CC administered systemically, particularly intravenously, typically at  
CC 0.2-200 mg, given 3 times over 7 days. Truncated TF, and its variants,  
CC localise specifically in tumour-associated blood vessels after systemic  
CC administration, even though they contain no targeting agent. They cause  
CC little if any injury to normal tissue; may produce a synergistic response  
CC when used with other antitumour agents and they eliminate the multi-step,  
CC and expensive, preparation of antibody-based targeting constructs. The  
CC present sequence encodes human TF, from the present invention.  
CC Sequence 13865 BP; 3711 A; 2955 C; 3240 G; 3959 T;  
SQ

Query Match 1.8%; Score 21; DB 1; Length 13865;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1153 ttccatctcaaaaaa 1173  
Db 11227 TTCCATCTCAAAAAA 11247

RESULT 14  
X37084/C  
ID X37084 standard; DNA; 16891 BP.  
AC X37084;  
DC 06-JUL-1999 (first entry)  
DE MEV gene sequence associated with familial mediterranean fever (FMF).  
KW MEV; protein pyrin; Familial Mediterranean Fever; FMF; human;  
KW FMF-associated mutant; hereditary disease; colchicine; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 1011..14450  
FT /tag= a  
FT /note= "contains introns"  
FT exon 1011..1287  
FT /tag= b  
FT /number= 1  
FT 1288..2806  
FT /tag= c  
FT /number= 1  
FT 2807..3439  
FT /tag= d  
FT intron  
FT exon

FT intron  
FT /number= 2  
FT 3440..7812  
FT /tag= e  
FT /number= 2  
FT 7813..8162  
FT /tag= f  
FT /number= 3  
FT 8163..8591  
FT /tag= g  
FT /number= 3  
FT 8592..8684  
FT /tag= h  
FT /number= 4  
FT 8685..10346  
FT /tag= i  
FT /number= 4  
FT 10347..10577  
FT /tag= j  
FT /number= 5  
FT 10578..11044  
FT /tag= k  
FT /number= 5  
FT 11045..11067  
FT /tag= l  
FT /number= 6  
FT 11068..13002  
FT /tag= m  
FT /number= 6  
FT 13003..13118  
FT /tag= n  
FT /number= 7  
FT 13119..13304  
FT /tag= o  
FT /number= 7  
FT 13305..13337  
FT /tag= p  
FT /number= 8  
FT 13338..13700  
FT /tag= q  
FT /number= 8  
FT /cons\_splice= (5'site= Yes, 3' site=No)  
FT 13701..13731  
FT /tag= r  
FT /number= 9  
FT 13732..13896  
FT /tag= s  
FT /number= 9  
FT 13897..14450  
FT /tag= t  
FT /number= 10  
PN WO9909169-A1.  
PD 25-FEB-1999.  
PF 20-AUG-1998; U17255.  
PR 21-AUG-1997; US-056217.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PI Aksentijevich I, Blake T, Centola M, Collins FS,  
PI Deng Z, Doggett NA, Fischel-Ghodsian N, Gumucio DL,  
PI Kastner DL, Liu PP, Pras M, Richards RI, Ricke DO,  
PI Sood R;  
DR WPI: 99-243555/20.  
DR P-PSDB: Y09001.  
DR A new gene, MEV, associated with Mediterranean Familial Fever  
PT Claim 6; Fig 1; 49pp; English.  
PS The present sequence represents a novel genomic nucleic acid sequence  
CC (MEV) encoding the protein pyrin associated with Familial Mediterranean  
CC acid encoding pyrin or its FMF-associated mutant, operably linked to a  
CC functional promoter, are used for the recombinant expression of the  
CC protein. The invention provides a method for diagnosing risk of FMF that  
CC comprises analyzing a patient sample for an amino acid or nucleic acid  
CC sequence of pyrin; and (b) correlating detection of a mutated sequence  
CC with risk of developing FMF; Diagnostic kits developed in this invention  
CC are used to identify and treat individuals at risk from FMF, a hereditary

CC disease prevalent in persons having a non-Ashkenazi Jewish, Armenian, Arab, or Turkish background. Prior art FMF treatment with colchicine is not effective in patients who are colchicine-resistant, and this invention will cover all individuals.

SQ Sequence 16891 BP; 3905 A; 4308 C; 4316 G; 4356 T;

Job time: 18912 sec

Query Match 1.8%; Score 21; DB 1; Length 16891;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1153 ttccatctcaaaaaaaaaa 1173

Db 9667 TTCCATCTCAAAAAAAAAA 9647

## RESULT 15

V57926/C  
ID V57926 standard; DNA; 235033 BP.

AC V57926;

DT 23-DEC-1998 (first entry)

DE Hereditary haemochromatosis subregion from an unaffected individual.

KW Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;

KW diagnosis; iron metabolism; NPT3; NPT4; Roret; BTF1; BTF2; BTF3;

KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;

KW type 1 sodium transport gene; ss.

OS Homo sapiens.

PN WO9814466-A1.

PD 09-APR-1998.

PF 30-SEP-1997; U17658.

PR 07-MAY-1997; US-852495.

PR 01-OCT-1996; US-724394.

PA (PROG-) PROGENITOR INC.

PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,

PI Tsuchihashi Z, Wolff RK;

DR WPI; 98-240014/21.

PT Hereditary haemochromatosis gene products - used to develop products

PT for the diagnosis and treatment of hereditary disorders in iron

PT metabolism

PS Example 2; Fig 8; 209pp; English.

CC The present invention describes hereditary haemochromatosis gene

CC products from the human haemochromatosis gene. The present sequence

CC represents a hereditary haemochromatosis subregion from an individual

CC unaffected by hereditary haemochromatosis (HH). Also described is a

CC method to determine the presence or absence of the common hereditary

CC haemochromatosis (HFE) gene mutation in an individual comprising:

CC (a) providing DNA or RNA from the individual; and (b) assessing the

CC DNA or RNA for the presence or absence of a haplotype or genotype where

CC the presence or absence of the haplotype genotype indicates the likely

CC presence of the HFE gene mutation in the genome of the individual. The

CC HFE gene sequences from the present invention can be used to develop

CC products for use in the diagnosis and treatment of HFE. The present

CC invention also describes BTF genes, which are homologues of the milk

CC protein butyrophilin (BT), and can be used in the production of agonists

CC and antagonists of BT function. Also described are: (1) a RoRet gene

CC which can be used to develop products for the study, diagnosis and

CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes

CC which are homologues of a type 1 sodium transport gene, and can

CC similarly be used for hypophosphatemia.

SQ Sequence 235033 BP; 68786 A; 48466 C; 49441 G; 68340 T;

Query Match

Best Local Similarity 1.8%; Score 21; DB 1; Length 235033;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1153 ttccatctcaaaaaaaaaa 1173

Db 95125 TTCCATCTCAAAAAAAAAA 95105

Search completed: May 1, 2000, 18:53:48





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:04:55 ; Search time 181.17 Seconds  
(without alignments)  
775.157 Million cell updates/sec

Title: US-09-215-435-117  
Perfect score: 1173  
Sequence: 1 gactgcttattggacacg.....tccatctcaaaaaaaaaa 1173

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 214294 seqs, 59861574 residues

Word size : 0

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2.6/ptodata/1/ina/5C\_COMB.seq.\*  
4: /cgn2.6/ptodata/1/ina/5D\_COMB.seq.\*  
5: /cgn2.6/ptodata/1/ina/6\_COMB.seq.\*  
6: /cgn2.6/ptodata/1/ina/PTUS9\_COMB.seq.\*  
7: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1111	94.7	1380	4 US-08-933-750C-74	Sequence 74, Appl
2	21	1.8	302	4 US-08-849-701-3	Sequence 3, Appl
3	21	1.8	1664	1 US-08-250-740-34	Sequence 34, Appl
4	21	1.8	1664	2 US-07-695-472B-3	Sequence 3, Appl
5	21	1.8	4741	2 US-07-695-472B-4	Sequence 4, Appl
6	21	1.8	4742	1 US-08-250-740-35	Sequence 35, Appl
7	21	1.8	24979	4 US-08-147-777-3	Sequence 3, Appl
8	21	1.8	24979	6 PCT-US93-03985-3	Sequence 3, Appl
9	21	1.8	246240	3 US-08-724-394A-20	Sequence 20, Appl
10	21	1.8	246240	3 US-08-724-394A-21	Sequence 21, Appl
11	21	1.8	246240	3 US-08-724-394A-22	Sequence 22, Appl
12	20	1.7	280	4 US-08-849-701-7	Sequence 7, Appl
13	20	1.7	301	3 US-08-332-766A-23	Sequence 23, Appl
14	20	1.7	807	3 US-08-531-927B-9	Sequence 9, Appl
15	20	1.7	1513	4 US-08-812-204-1	Sequence 1, Appl
16	20	1.7	1706	1 US-07-906-871-11	Sequence 11, Appl
17	20	1.7	1776	3 US-08-531-927B-1	Sequence 1, Appl
18	20	1.7	1875	3 US-08-683-743-3	Sequence 3, Appl
19	20	1.7	1901	6 PCT-US93-05000-32	Sequence 32, Appl
20	20	1.7	2022	3 US-08-464-517-32	Sequence 32, Appl
21	20	1.7	2022	4 US-08-246-361A-32	Sequence 32, Appl
22	20	1.7	2048	1 US-07-602-608-11	Sequence 11, Appl
23	20	1.7	2048	1 US-08-261-578-11	Sequence 11, Appl
24	20	1.7	2115	2 US-08-395-800A-7	Sequence 7, Appl
25	20	1.7	2679	2 US-08-479-328-1	Sequence 1, Appl
26	20	1.7	2679	2 US-08-761-119-1	Sequence 1, Appl

Sequence 1, Appl  
Sequence 1, Appl  
Sequence 55, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 67, Appl  
Sequence 8, Appl  
Sequence 3, Appl  
Sequence 12, Appl  
Sequence 49, Appl  
Sequence 49, Appl  
Sequence 53, Appl  
Sequence 53, Appl  
Sequence 12, Appl  
Patent No. 5196523  
Sequence 4, Appl  
Sequence 4, Appl

27 20 1.7 2679 3 US-08-668-128B-1  
28 20 1.7 2679 3 US-08-905-445-1  
29 20 1.7 2886 4 US-08-687-080-55  
30 20 1.7 2923 1 US-08-480-449-1  
31 20 1.7 2923 4 US-08-660-542-1  
32 20 1.7 2923 4 US-08-989-107-1  
33 20 1.7 3507 2 US-08-832-883-67  
34 20 1.7 3507 3 US-08-832-877-67  
35 20 1.7 3607 3 US-08-629-001A-8  
36 20 1.7 4080 3 US-08-710-249-8  
37 20 1.7 4326 3 US-08-852-807-12  
38 20 1.7 4576 2 US-08-832-883-49  
39 20 1.7 4576 3 US-08-832-877-49  
40 20 1.7 4586 2 US-08-832-883-53  
41 20 1.7 4586 3 US-08-832-877-53  
42 20 1.7 5470 2 US-08-441-139-12  
43 20 1.7 5470 7 5196523-5  
44 20 1.7 6063 1 US-08-195-744-4  
45 20 1.7 6063 3 US-08-788-279-4

## ALIGNMENTS

RESULT 1  
US-08-933-750C-74  
; Sequence 74, Application US/08933750C  
; Patent No. 5932442  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Bandman, Olga  
; APPLICANT: Shah, Purvi  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Yue, Henry  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
; NUMBER OF SEQUENCES: 98  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/933,750C  
; FILING DATE: September 23, 1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0356 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 74:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1380 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:

; LIBRARY: LKODNOT03  
 ; CLONE: 1572888  
 US-08-933-750C-74

```
Query Match          94.7%; Score 1111; DB 4; Length 1380;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1161: Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Qy	1	gagctcttatgggacacccggttctctgcggcgctcttaacgtgtctgtcgcgcgcga	60
Db	46	GAGCTGCTTATGGGACACCGCTTCTCGCGGGCTCTTAACGCTGCTGCTGCCCGCCCA	105
Qy	61	ccctgtataccggcagccgcatgctcgttcagagtcctgcgcgcgcgcgcgcgcgcgc	120
Db	106	CCCCTGATACCGGCACCGCATGCTCGGTTCAGAGTCGCTCCGCGCCCCCAACCATCC	165
Qy	121	cgagcaaaactcatggcacccgcgcgaatcgggacgcacaaatggcaccttccactcgac	180
Db	166	CGCAGCAAACTCATGTGCACCGCCCGAATCGGGACGACAAATGCGACTTCCACTCGGAC	225
Qy	181	gaggcactggcatgcgactgttccgctccttcgcggagtaccgggagtcagagattgtg	240
Db	226	GAGGCACTGGCATGCGACTGCTTCCGCTCCTCGCGGAGTACCGGGATGCAGAGATTGTG	285
Qy	241	cgagaccgggatcccgaaaaactcgtctcctgtgacatcgtgtgagcgtggtggggcgag	300
Db	286	CGGACCGGGATCCCCAAAATTCGCTTCTGTGACATCGGTGTGAGCTGGGGGGCGAG	345
Qy	301	tacgacctcgagacacccgatataccatcacagagtgctttccacagagaccatgagc	360
Db	346	TACGACCTCGGAGACACCGATATGACCATCACCAGAGGTCTTTCCACAGAGACCATGAGC	405
Qy	361	tcctgtccccctgggagccgtggcagaccaagctgagcagtgcgggactcatctatctg	420
Db	406	TCCTGTCCCTCGGAAGCGGTGGCAGACCAAGCTGAGCAGTGGGAGAGTGGATGCTGTGCAAT	465
Qy	421	cacttggggcacaaagctgtgccagtgctgggcactagtgagaggacacatcggtg	480
Db	466	CAC TTCGGGCACAAAGTGTCTGGCCCCAGTTGCTTGGCCTATGTGAAGAGGACGACATGGTG	525
Qy	481	ggcacctctatgacaagatgtatgagaactttgggagggtggatgctgtggacaat	540
Db	526	GGCACCTCTATGACAAGATGTATGAGAACTTTGTGGAGGAGTGGATGCTGTGGCAAT	585
Qy	541	gggatctccagtgggcagaggggagcctcgatgcactgacacactaccctgagtgca	600
Db	586	GGGATATCCCAAGTGGGACAGGGGGAGCCTCGATATGACATGACCTGACCATACCCCTGAGTGCA	645
Qy	601	cgagtgtcgcacttaactcctactggaaccaccccgaccgaagacactgaggcagggttc	660
Db	646	CGAGTGTCTCGACTTAATCTACTTGGAAACCACCCCGACCAAGACACTGAGCGAGGGTTC	705
Qy	661	aagcgtgcaatgatctggttcaagaggagttcttcagagatagattttaccacaac	720
Db	706	AAGCGTGCAATGGATCTGTTCAAGAGGAGTTCTGCAGAGATTAGATTTCTACCAACAC	765
Qy	721	agctggctgcagccgggcttgggtgggaaggcccttgcgcagcgattccaggtggac	780
Db	766	AGCTTGGCTGCCAGCCCGGGCTTTGGTGGAAAGAGGCCCTTGCCACGAGTTCACAGGTGAC	825
Qy	781	ccaagtgagagatatgtggaactgcggaaggtgcatgtccctgggaaggagcatctctac	840
Db	826	CCAAGTGAGAGATTGTGGAAC TGGGAAAGGTGATGTCTCCCTGGGAAGGACATCTCTAC	885
Qy	841	cacctgaaactggctgtccctccagtggccatctcttggatttatcacactgaccag	900
Db	886	CACCTGGAACTCGGGCTGTCCCTCCAGTGGCCACTTCTTTGTTATCTACACTGACCAG	945
Qy	901	gctggacagtggcgaatacacagttgtgcccgaaggagccccactcatctcaagcggctg	960
Db	946	GCTGGACAGTGGCAATACAGTGTGTGCCAAGAGGCCCACTCATTTCAAAGCGGGTG	1005

Qy	961	cccttcgccagagccatcgcgcggtcttcgggacagagccctcgacacaggtcagtgggatc	1023
Db	1006	CCCTGCCAGAGCATGGCGGGGTCTCGGACGAGGCCCTGGACCAGGTCTAGTGGGATC	1065
Qy	1021	ccctggctgcatcttgctccatcaagcggcttcattggcggtaccgcaccccgagaggggt	1080
Db	1066	CCTGGCTGCACTCTGCTCCATGCAAGCGGCTTCATTGGCGGTACCCGACCCGAGAGGGT	1125
Qy	1081	gccttgagcatggcccggtgccaccttggccagcgcctcatcctcccaaaatctcctag	1140
Db	1126	GCCTTGAGCATGCGCCGTGCCACCTTGGCCAGCGCTCATACCTCCCACAAAATCTCTAG	1185
Qy	1141	tctaataaaaccttccatctca	1162
Db	1186	TCTAATAAAACCTTCCATCTCA	1207

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

CLASSIFICATION: 43P  
PRIOR APPLICATION DATA: PCT/JP95/02734  
APPLICATION NUMBER: 27-DEC-1995  
FILING DATE: 27-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: EIKEN1.001APC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 302 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: Alu sequence BLUR2  
US-08-849-701-3

Query Match 1.8%; Score 21; DB 4; Length 302;  
Best Local Similarity 100.0%; Pred. No. 0.7;  
Matches 21: Conservative 0; Mismatches 0; Indels

Qy 1153 ttccatctcaaaaaaaaaa 1173  
|||||  
Db 272 TTCCATCTCAAAAAAAAAA 292

```
RESULT 3
US-08-250-740-34/c
; Sequence 34, Application US/08250740
; Patent No. 5686240
; GENERAL INFORMATION:
; APPLICANT: Schuchman, Edward H.
; APPLICANT: Desnick, Robert J.
; TITLE OF INVENTION: Acid Sphingomyelinase Gene and Diagnosis
; TITLE OF INVENTION: of Niemann-Pick Disease
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/250,740
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 6923-038
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1664 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-250-740-34
```

```
Query Match 1.8%; Score 21; DB 1; Length 1664;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1153 ttccatctcaaaaaaaaaa 1173
|||||
Db 498 TTCCATCTCAAAAAAAAAA 478
```

```
RESULT 4
US-07-695-472B-3/c
; Sequence 3, Application US/07695472B
; Patent No. 5773278
; GENERAL INFORMATION:
; APPLICANT: Schuchman, Edward H.
; APPLICANT: Desnick, Robert J.
; TITLE OF INVENTION: The Acid Sphingomyelinase Gene and
; TITLE OF INVENTION: Diagnosis of Niemann-Pick Disease
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/695,472B
; FILING DATE: 19910503
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6923-014
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 7908864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1664 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-07-695-472B-3

Query Match 1.8%; Score 21; DB 2; Length 1664;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1153 ttccatctcaaaaaaaaaa 1173
|||||
Db 498 TTCCATCTCAAAAAAAAAA 478

RESULT 5
US-07-695-472B-4/c
; Sequence 4, Application US/07695472B
; Patent No. 5773278
; GENERAL INFORMATION:
; APPLICANT: Schuchman, Edward H.
; APPLICANT: Desnick, Robert J.
; TITLE OF INVENTION: The Acid Sphingomyelinase Gene and
; TITLE OF INVENTION: Diagnosis of Niemann-Pick Disease
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/695,472B
; FILING DATE: 19910503
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6923-014
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 7908864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1664 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
```

7

```
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03985
; FILING DATE: 19930428
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5478
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5325
; TELEFAX: 713/651-5246
; TELEX: 762829
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24979 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US93-03985-3

Query Match 1.8%; Score 21; DB 6; Length 24979;
Best Local Similarity 100.0%; Pred. No. 0.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1153 ttccatctcaaaaaa 1173
|||||
DB 2469 TTCCATCTCAAAAAA 2449

RESULT 9
US-08-724-394A-20
; Sequence 20, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0300
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
```

```
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-20

Query Match 1.8%; Score 21; DB 3; Length 246240;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1153 ttccatctcaaaaaa 1173
|||||
DB 144325 TTCCATCTCAAAAAA 144345

RESULT 10
US-08-724-394A-21
; Sequence 21, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
```

NAME/KEY: misc\_feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H. CONTIG"  
US-08-724-394A-21

Query Match 1.8%; Score 21; DB 3; Length 246240;  
Best Local Similarity 100.0%; Pred. No. 0.56;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1153 ttccatctcaaaaaaaaaa 1173  
|||||  
Db 144325 TTCCATCTCAAAAAAAAAA 144345

RESULT 11  
US-08-724-394A-22  
; Sequence 22, Application US/08724394A  
; Patent No. 5872237  
; GENERAL INFORMATION:  
; APPLICANT: Feder, John N.  
; APPLICANT: Kronmal, Gregory S.  
; APPLICANT: Lauer, Peter M.  
; APPLICANT: Ruddy, David A.  
; APPLICANT: Thomas, Winston  
; APPLICANT: Tsuchihashi, Zenta  
; APPLICANT: Wolff, Roger K.  
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/724,394A  
; FILING DATE: 01-OCT-1996  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fitts, Renee A.  
; REGISTRATION NUMBER: 35,136  
; REFERENCE/DOCKET NUMBER: 017957-000100  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-576-0200  
; TELEFAX: 415-576-0300  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 246240 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: cDNA  
; FEATURE:

Query Match 1.8%; Score 21; DB 3; Length 246240;  
Best Local Similarity 100.0%; Pred. No. 0.56;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1153 ttccatctcaaaaaaaaaa 1173  
|||||

Db 144325 TTCCATCTCAAAAAAAAAA 144345

RESULT 12  
US-08-849-701-7  
; Sequence 7, Application US/08849701  
; Patent No. 5922544  
; GENERAL INFORMATION:  
; APPLICANT: Miyai, Kiyoshi  
; APPLICANT: Naichou, Tsutomu  
; APPLICANT: Yonekawa, Toshihiro  
; TITLE OF INVENTION: Method of cell Detection  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobb, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/849,701  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP95/02734  
; FILING DATE: 27-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Altman, Daniel E  
; REGISTRATION NUMBER: 34,115  
; REFERENCE/DOCKET NUMBER: EIKEN1.001APC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 714-760-0404  
; TELEFAX: 714-760-9502  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 280 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: Alu sequence BLUR11

Query Match 1.7%; Score 20; DB 4; Length 280;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1154 ttccatctcaaaaaaaaaa 1173  
|||||  
Db 254 TTCCATCTCAAAAAAAAAA 273

RESULT 13  
US-08-332-766A-23  
; Sequence 23, Application US/08332766A  
; Patent No. 5843647  
; GENERAL INFORMATION:  
; APPLICANT: JEFFREYS, Alec J.  
; APPLICANT: ARMOUR, John  
; TITLE OF INVENTION: SIMPLE TANDEM REPEATS  
; NUMBER OF SEQUENCES: 125  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DABBY & CUSHMAN, L.L.P.  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington

STATE: D. C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: IBM PC compatible  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/332,766A  
FILING DATE: 01-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9326052.9  
FILING DATE: 21-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BIRD, Donald J.  
REGISTRATION NUMBER: 25,323  
REFERENCE/DOCKET NUMBER: 217211/M94/0434/CB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 301 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-332-766A-23

Query Match 1.7%; Score 20; DB 3; Length 301;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1154 tccatctcaaaaaaaaaa 1173  
|||||  
DB 42 TCCATCTCAAAAAAAAAA 61

RESULT 14  
US-08-531-927B-9  
Sequence 9, Application US/08531927B  
Patent No. 5840491  
GENERAL INFORMATION:  
APPLICANT: Kakizuka, Akira  
TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph  
Patent No. 5840491  
TITLE OF INVENTION: Disease Gene and Uses Thereof  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173-4799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/531,927B  
FILING DATE: 21-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP H6-251600  
FILING DATE: 21-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: ATH95-01A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 807 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-531-927B-9

Query Match 1.7%; Score 20; DB 3; Length 807;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1154 tccatctcaaaaaaaaaa 1173  
|||||  
DB 750 TCCATCTCAAAAAAAAAA 769

RESULT 15  
US-08-812-204-1/c  
Sequence 1, Application US/08812204  
Patent No. 5965790  
GENERAL INFORMATION:  
APPLICANT: Acton, Susan L.  
TITLE OF INVENTION: SR-BI REGULATORY SEQUENCES AND  
METHODS OF INVENTION: THERAPEUTIC METHODS OF USE  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/812,204  
FILING DATE: 06-MAR-1997  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: MIA-014.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1613 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-812-204-1

Query Match 1.7%; Score 20; DB 4; Length 1613;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1154 tccatctcaaaaaaaaaa 1173  
|||||  
DB 385 TCCATCTCAAAAAAAAAA 366

Search completed: May 1, 2000, 16:10:59  
Job time: 18724 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 19:59:52 ; Search time 4088.29 Seconds  
(without alignments)  
-983.901 Million cell updates/sec

Title: US-09-215-435-117  
Perfect score: 1173  
Sequence: 1 gagtgcttatggacacg.....tccatctcaaaaaaaaaa 1173

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 5265378 seqs, -1714608768 residues

Word size : 0

Total number of hits satisfying chosen parameters: 10530736

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Pending Patents.NA.\*

- 1: /cgn2\_6/ptodata/1/pna/PCTUS9\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/pna/US081A\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pna/US081B\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/pna/US081C\_COMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pna/US081D\_COMB.seq.\*
- 8: /cgn2\_6/ptodata/1/pna/US082A\_COMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pna/US082B\_COMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pna/US082C\_COMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pna/US083A\_COMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pna/US083B\_COMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pna/US084A\_COMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pna/US084B\_COMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pna/US084C\_COMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pna/US087A\_COMB.seq.\*
- 19: /cgn2\_6/ptodata/1/pna/US087B\_COMB.seq.\*
- 20: /cgn2\_6/ptodata/1/pna/US087C\_COMB.seq.\*
- 21: /cgn2\_6/ptodata/1/pna/US088A\_COMB.seq.\*
- 22: /cgn2\_6/ptodata/1/pna/US088B\_COMB.seq.\*
- 23: /cgn2\_6/ptodata/1/pna/US088C\_COMB.seq.\*
- 24: /cgn2\_6/ptodata/1/pna/US088D\_COMB.seq.\*
- 25: /cgn2\_6/ptodata/1/pna/US089A\_COMB.seq.\*
- 26: /cgn2\_6/ptodata/1/pna/US089B\_COMB.seq.\*
- 27: /cgn2\_6/ptodata/1/pna/US089C\_COMB.seq.\*
- 28: /cgn2\_6/ptodata/1/pna/US089D\_COMB.seq.\*
- 29: /cgn2\_6/ptodata/1/pna/US089E\_COMB.seq.\*
- 30: /cgn2\_6/ptodata/1/pna/US089F\_COMB.seq.\*
- 31: /cgn2\_6/ptodata/1/pna/US089G\_COMB.seq.\*
- 32: /cgn2\_6/ptodata/1/pna/US089H\_COMB.seq.\*
- 33: /cgn2\_6/ptodata/1/pna/US089I\_COMB.seq.\*
- 34: /cgn2\_6/ptodata/1/pna/US090A\_COMB.seq.\*
- 35: /cgn2\_6/ptodata/1/pna/US090B\_COMB.seq.\*
- 36: /cgn2\_6/ptodata/1/pna/US090C\_COMB.seq.\*
- 37: /cgn2\_6/ptodata/1/pna/US090D\_COMB.seq.\*
- 38: /cgn2\_6/ptodata/1/pna/US090E\_COMB.seq.\*
- 39: /cgn2\_6/ptodata/1/pna/US091A\_COMB.seq.\*
- 40: /cgn2\_6/ptodata/1/pna/US091B\_COMB.seq.\*
- 41: /cgn2\_6/ptodata/1/pna/US091C\_COMB.seq.\*
- 42: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq.\*
- 43: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq.\*

- 44: /cgn2\_6/ptodata/1/pna/US092C\_COMB.seq.\*
- 45: /cgn2\_6/ptodata/1/pna/US092D\_COMB.seq.\*
- 46: /cgn2\_6/ptodata/1/pna/US092E\_COMB.seq.\*
- 47: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq.\*
- 48: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq.\*
- 49: /cgn2\_6/ptodata/1/pna/US093C\_COMB.seq.\*
- 50: /cgn2\_6/ptodata/1/pna/US093D\_COMB.seq.\*
- 51: /cgn2\_6/ptodata/1/pna/US093E\_COMB.seq.\*
- 52: /cgn2\_6/ptodata/1/pna/US093F\_COMB.seq.\*
- 53: /cgn2\_6/ptodata/1/pna/US094A\_COMB.seq.\*
- 54: /cgn2\_6/ptodata/1/pna/US094B\_COMB.seq.\*
- 55: /cgn2\_6/ptodata/1/pna/US094C\_COMB.seq.\*
- 56: /cgn2\_6/ptodata/1/pna/US094D\_COMB.seq.\*
- 57: /cgn2\_6/ptodata/1/pna/US094E\_COMB.seq.\*
- 58: /cgn2\_6/ptodata/1/pna/US094F\_COMB.seq.\*
- 59: /cgn2\_6/ptodata/1/pna/US094G\_COMB.seq.\*
- 60: /cgn2\_6/ptodata/1/pna/US094H\_COMB.seq.\*
- 61: /cgn2\_6/ptodata/1/pna/US094I\_COMB.seq.\*
- 62: /cgn2\_6/ptodata/1/pna/US094J\_COMB.seq.\*
- 63: /cgn2\_6/ptodata/1/pna/US094K\_COMB.seq.\*
- 64: /cgn2\_6/ptodata/1/pna/US094L\_COMB.seq.\*
- 65: /cgn2\_6/ptodata/1/pna/US094M\_COMB.seq.\*
- 66: /cgn2\_6/ptodata/1/pna/US094N\_COMB.seq.\*
- 67: /cgn2\_6/ptodata/1/pna/US094O\_COMB.seq.\*
- 68: /cgn2\_6/ptodata/1/pna/US094P\_COMB.seq.\*
- 69: /cgn2\_6/ptodata/1/pna/US094Q\_COMB.seq.\*
- 70: /cgn2\_6/ptodata/1/pna/US094R\_COMB.seq.\*
- 71: /cgn2\_6/ptodata/1/pna/US094S\_COMB.seq.\*
- 72: /cgn2\_6/ptodata/1/pna/US094T\_COMB.seq.\*
- 73: /cgn2\_6/ptodata/1/pna/US094U\_COMB.seq.\*
- 74: /cgn2\_6/ptodata/1/pna/US094V\_COMB.seq.\*
- 75: /cgn2\_6/ptodata/1/pna/US094W\_COMB.seq.\*
- 76: /cgn2\_6/ptodata/1/pna/US094X\_COMB.seq.\*
- 77: /cgn2\_6/ptodata/1/pna/US094Y\_COMB.seq.\*
- 78: /cgn2\_6/ptodata/1/pna/US094Z\_COMB.seq.\*
- 79: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq.\*
- 80: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq.\*
- 81: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq.\*
- 82: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq.\*
- 83: /cgn2\_6/ptodata/1/pna/US095E\_COMB.seq.\*
- 84: /cgn2\_6/ptodata/1/pna/US095F\_COMB.seq.\*
- 85: /cgn2\_6/ptodata/1/pna/US095G\_COMB.seq.\*
- 86: /cgn2\_6/ptodata/1/pna/US095H\_COMB.seq.\*
- 87: /cgn2\_6/ptodata/1/pna/US095I\_COMB.seq.\*
- 88: /cgn2\_6/ptodata/1/pna/US095J\_COMB.seq.\*
- 89: /cgn2\_6/ptodata/1/pna/US095K\_COMB.seq.\*
- 90: /cgn2\_6/ptodata/1/pna/US095L\_COMB.seq.\*
- 91: /cgn2\_6/ptodata/1/pna/US095M\_COMB.seq.\*
- 92: /cgn2\_6/ptodata/1/pna/US095N\_COMB.seq.\*
- 93: /cgn2\_6/ptodata/1/pna/US095O\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1173	100.0	1173	42	US-09-215-435-117
2	1173	100.0	1173	69	US-60-081-563-74
3	1111	94.7	1380	27	US-08-933-750-74
4	1111	94.7	1380	43	US-09-234-613-74
5	1111	94.7	1402	54	US-09-474-436-8324
6	1111	94.7	1456	92	US-09-515-126-14066
7	991	84.5	1382	87	US-60-172-360-23018
8	773	65.9	1297	50	US-09-359-922-3400
9	773	65.9	1297	92	US-09-359-922-3400
10	652	55.6	1199	42	US-09-205-070-7647
11	652	55.6	1199	49	US-09-340-623-7647
12	437	37.3	566	50	US-09-353-690-11806
13	436	37.2	569	44	US-09-277-227-22988

14	399	34.0	446	50	US-09-353-690-7366	Sequence 7366, App
15	398	33.9	404	50	US-09-353-690-11003	Sequence 11003, App
16	385	32.8	516	46	US-09-293-972-5796	Sequence 5796, App
17	385	33.8	516	92	US-09-240-371-5150	Sequence 5150, App
18	376	32.1	386	44	US-09-277-227-16632	Sequence 16632, App
19	376	32.1	386	48	US-09-332-782-37472	Sequence 37472, App
20	376	32.1	532	51	US-09-362-510-48543	Sequence 48543, App
21	367	31.3	388	50	US-09-359-067-116	Sequence 116, App
22	361	30.8	406	46	US-09-293-972-29225	Sequence 29225, App
23	329	28.0	329	65	US-60-048-722-212	Sequence 212, App
24	324	27.6	376	48	US-09-332-782-20518	Sequence 20518, App
25	324	27.6	376	92	US-09-515-694-20518	Sequence 20518, App
26	324	27.6	421	50	US-09-489-036-40497	Sequence 40497, App
27	323	27.5	572	92	US-09-489-036-26803	Sequence 26803, App
28	320	27.3	422	28	US-08-824-029-256	Sequence 256, App
29	319	27.2	473	48	US-09-321-214-30273	Sequence 30273, App
30	319	27.2	473	92	US-09-516-335-30273	Sequence 30273, App
31	318	27.1	481	92	US-09-321-214-30272	Sequence 30272, App
32	318	27.1	481	92	US-09-516-335-30272	Sequence 30272, App
33	304	26.9	874	93	US-60-185-212-129	Sequence 129, App
34	304	25.9	304	65	US-60-048-722-2032	Sequence 2032, App
35	302	25.7	361	50	US-09-359-067-4404	Sequence 4404, App
36	296	25.2	395	53	US-09-431-511-9242	Sequence 9242, App
37	296	25.2	395	53	US-09-431-511-9302	Sequence 9302, App
38	292	24.9	388	92	US-09-489-036-2184	Sequence 2184, App
39	291	24.8	530	47	US-09-306-609-10548	Sequence 10548, App
40	291	24.8	530	92	US-09-322-251-10548	Sequence 10548, App
41	289	24.6	343	41	US-09-179-473-1136	Sequence 1136, App
42	289	24.6	343	48	US-09-328-351-1136	Sequence 1136, App
43	273	23.3	273	34	US-09-015-081-757	Sequence 757, App
44	273	23.3	273	62	US-60-036-169-757	Sequence 757, App
45	273	23.3	319	92	US-09-489-036-12752	Sequence 12752, App

## ALIGNMENTS

RESULT	1	
US-09-215-435-117		
Sequence 117, Application US/09215435A		
GENERAL INFORMATION:		
APPLICANT: Dumas Milne Edwards, Jean-Baptiste		
APPLICANT: Duclert Aymeric,		
APPLICANT: Bougueleret Lydie		
TITLE OF INVENTION: Extended cDNAs		
FILE REFERENCE: GENSET.019A		
CURRENT APPLICATION NUMBER: US/09/215,435A		
CURRENT FILING DATE: 1998-12-17		
EARLIER APPLICATION NUMBER: 60/069,957		
EARLIER FILING DATE: 1997-12-17		
EARLIER APPLICATION NUMBER: 60/074,121		
EARLIER FILING DATE: 1998-2-9		
EARLIER APPLICATION NUMBER: 60/081,563		
EARLIER FILING DATE: 1998-4-13		
EARLIER APPLICATION NUMBER: 60/096,116		
EARLIER FILING DATE: 1998-8-10		
NUMBER OF SEQ ID NOS: 519		
SOFTWARE: Patent.pm		
SEQ ID NO 117		
LENGTH: 1173		
TYPE: DNA		
ORGANISM: Homo sapiens		
FEATURE:		
NAME/KEY: CDS		
LOCATION: 10..1137		
FEATURE:		
NAME/KEY: sig_peptide		
LOCATION: 10..72		
OTHER INFORMATION: Von Heijne matrix		
OTHER INFORMATION: score 6.5		
OTHER INFORMATION: seq LITLLLPPPLPYT/RH		
FEATURE:		
NAME/KEY: polyA_signal		

```
Db 901 gctggacagtggcgaataacagtgtgtgccaaaggagcccccactcattccaaagccggctg 960
QY 961 cccctccagagacatggcgggtctctcgggacgagcccttgaccaggtcagtgsggac 1020
Db 961 cccctccagagacatggcgggtctctcgggacgagcccttgaccaggtcagtgsggac 1020
QY 1021 cctggctgatctctccatcgaagcggcttctcattggcgggtcaccgcaccgagaggg 1080
Db 1021 cctggctgatctctccatcgaagcggcttctcattggcgggtcaccgcaccgagaggg 1080
QY 1081 gccctgagcatgcccgctgcacacttggccagcgcgtctacatcccccacaaatctcctag 1140
Db 1081 gccctgagcatgcccgctgcacacttggccagcgcgtctacatcccccacaaatctcctag 1140
QY 1141 tctaataaaccttcctcctcctcaaaaaa 1173
Db 1141 tctaataaaccttcctcctcctcaaaaaa 1173

RESULT 2
US-60-081-563-74
; Sequence 74, Application US/60081563
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste Dumas Milne
; APPLICANT: Duclert, Aymeric
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SECRETED PROTEINS
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/081,563
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.027PR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1173 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..893
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 816..857
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 6.8
; OTHER INFORMATION: seq SLEGASLPPGIWA/VP
; FEATURE:
; NAME/KEY: polya_signal
; LOCATION: 1144..1149
; FEATURE:
```

```
; NAME/KEY: poly_a
; LOCATION: 1162..1173
; IDENTIFICATION METHOD: blastn
; FEATURE:
; NAME/KEY: est
; LOCATION: 755..1161
; IDENTIFICATION METHOD: blastn
; OTHER INFORMATION: identity 99
; OTHER INFORMATION: region 11..417
; OTHER INFORMATION: id AA401345
; FEATURE:
; NAME/KEY: est
; LOCATION: 565..849
; IDENTIFICATION METHOD: blastn
; OTHER INFORMATION: identity 99
; OTHER INFORMATION: region 6..290
; OTHER INFORMATION: id H64785
; FEATURE:
; NAME/KEY: est
; LOCATION: 846..940
; IDENTIFICATION METHOD: blastn
; OTHER INFORMATION: identity 92
; OTHER INFORMATION: region 288..382
; OTHER INFORMATION: id H64785
; FEATURE:
; NAME/KEY: est
; LOCATION: 726..1057
; IDENTIFICATION METHOD: blastn
; OTHER INFORMATION: identity 100
; OTHER INFORMATION: region 80..411
; OTHER INFORMATION: id AA084391
; FEATURE:
; NAME/KEY: est
; LOCATION: 1056..1137
; IDENTIFICATION METHOD: blastn
; OTHER INFORMATION: identity 95
; OTHER INFORMATION: region 1..82
; OTHER INFORMATION: id AA084391
; FEATURE:
; NAME/KEY: est
; LOCATION: 143..404
; IDENTIFICATION METHOD: blastn
; OTHER INFORMATION: identity 99
; OTHER INFORMATION: region 160..421
; OTHER INFORMATION: id AA401585
; FEATURE:
; NAME/KEY: est
; LOCATION: 1..144
; IDENTIFICATION METHOD: blastn
; OTHER INFORMATION: identity 100
; OTHER INFORMATION: region 19..162
; OTHER INFORMATION: id AA401585
; FEATURE:
; NAME/KEY: est
; LOCATION: 897..1071
; IDENTIFICATION METHOD: blastn
; OTHER INFORMATION: identity 95
; OTHER INFORMATION: region 96..270
; OTHER INFORMATION: id H50235
; FEATURE:
; NAME/KEY: est
; LOCATION: 1056..1161
; IDENTIFICATION METHOD: blastn
; OTHER INFORMATION: identity 91
; OTHER INFORMATION: region 7..112
; OTHER INFORMATION: id H50235
; FEATURE:
; NAME/KEY: est
; LOCATION: 843..898
; IDENTIFICATION METHOD: blastn
; OTHER INFORMATION: identity 100
; OTHER INFORMATION: region 270..325
; OTHER INFORMATION: id H50235
```





```
Db 1186 TCTAATAAAACCTTCCATCTCA 1207
RESULT 4
US-09-234-613-74
; Sequence 74, Application US/09234613
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,613
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1380 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LNDNOT03
; CLONE: 1572886
US-09-234-613-74

Query Match 94.7%; Score 1111; DB 43; Length 1380;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1161; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gagctgcttatgggacacgcgtctctcgcgccgtctttaacgctgctgctgccgcgcga 60
Db 46 GAGCTGCTTATGGGACACCGCTTCTGCGCGCCCTTTAAAGCTGCTGCTGCGCGCGCA 105
Qy 61 cccctgtatacccgacccgcgatgctgctgctccagtgctccgcccccaaacgatcc 120
Db 106 CCCCTGTATACCGGACCGCATGCTCGGTCCAGATCCGTCGCGCCGCCCAACGATCC 165
Qy 121 cgcagcaactatggaccccccgaatcgggagcacaatggcacattccactcgagac 180
Db 166 CCACCAAACTATGGGACCCCGGAAATGGGACGCAATGGCACCCTTCCACTGGCAC 225
Qy 181 gaggcactggcagtcgcactgcttgcctcctcgccggagtagccggagtagcagagattgtg 240

Db 226 GAGGCACTGGCATGCGCACTGCTTCGCTCTCGCGAGTACCGGATGACAGATTGTG 285
Qy 241 cggaccgggaccccgaaacactcgcttctgtgacatcgatgacgctggggggcgag 300
Db 286 CGGACCCGGGATCCCGAAACACTCGCTCTGTGACATCGTGTGGGACGTTGGGGGGCGAG 345
Qy 301 taegaccctcggagacaccgcgatgatgaccatcacagaggctctttcacagagaccatgagc 360
Db 346 TACGACCCTCGGAGACACCGATATGACCATCACAGAGGTCTTTTCACAGAGACCATGAGC 405
Qy 361 tccctgtccctcggggagccgctggcgagacaaagctgagcagtgccgggactcatctctg 420
Db 406 TCCCTGTCCCTGGGAAGCCGTGGCAGACCAAGCTGACAGTGGCGGACTCATCTATCTG 465
Qy 421 cactcggggcacaagctgctggcccgcttctgctggcactagtgaagagacacagcatgtg 480
Db 466 CACTTCGGGCACAAGCTGCTGGCCCGAGTTGCTGGGCACTAGTGAAGAGGACAGCATGGTG 525
Qy 481 ggcaccctctatgacaagatgtatgagaactttgtgagggaggtggatgctgtggacaat 540
Db 526 GGCACCCTCTATGACAAGATGTATGAGAACTTTGTGGAGGAGGTGGATGCTGTGGACAAT 585
Qy 541 gggatctcccagtgggcagagggggagccctcgatatactgactgaccactaccctgagtgca 600
Db 586 GGGATCTCCAGTGGGAGAGGGGGAGCCCTCGATATGCACTGACCTACCTAGCCCTGAGTGCA 645
Qy 601 cgagtgtcgtgacttaactcctgaacacaccccccacacacacacacacacacacacacac 660
Db 646 CGAGTTGCTGACTTAATCTCTCTGGAACACCCCGACCAAGACACTGAGGCAAGGGTTC 705
Qy 661 aagcgtgcaatggatctggttcaagagaggtttctgcagagattagatttctacacacac 720
Db 706 AAGCGTGCAATGGATCTGTTCAAGAGAGAGTTCCTGTCAGAGATTAGATTCTTACCAACAC 765
Qy 721 agctggctgcagcccgccctggtggaagagcccttcccagcgatccaggtggagac 780
Db 766 AGCTGGCTGCCAGCCCGGGCCCTTGGTGGAAAGAGGCCCTTGCCAGCGCATTCACAGGTGGAC 825
Qy 781 ccaagtggagagattgtggaactggcgaaggtgcatgtccctgggaagagcatctctac 840
Db 826 CCAAGTGGAGAGATTGTGGAACGGCGAAAGGTGCATGTCCCTGGAAAGGAGCATCTCTAC 885
Qy 841 cactgggaatctgggctgtccctccagtggtccatctcttcttcttcttcttcttcttcttct 900
Db 886 CACCTGGAATCTGGGCTGTCCCTCCAGTGGCCATCTCTTGTGTTATCTACTGACACAG 945
Qy 901 gctggacagtgggcgaatacacagtgtgtgcccaagagagcccccactcattccaaagccgctg 960
Db 946 GCTGGACAGTGGCGAATACAGTGTGTGCCCCAAGGAGCCCCCACTCATTCCAAAGCCGCGTG 1005
Qy 961 cccctgcagagccatggcgggggtcttcggggagagagccctgagaccaggtcagtgggatc 1020
Db 1006 CCCCTGCCAGAGCCATGGCGGGGTCTTCGGGACGAGGCCCTTGACACAGGTTCAGTGGGATC 1065
Qy 1021 cctggtgcattcttgcctgaagcggttcttcttcttcttcttcttcttcttcttcttcttct 1080
Db 1066 CCTGGCTGCATCTTCGTCCATGCAAGCGGCTTCATTGGCGGTTCACCGCACCCGAGAGGGT 1125
Qy 1081 gccttgagcatggcccgctgcacacttgccagcgctcatcacctcccacaaatctctctag 1140
Db 1126 GCCTTGAGCATGGCCCGTGCCACCTTGGCCCGACGCTTATGGCGGTTCACCGCACCCGAGAGGGT 1185
Qy 1141 tctaataaaacaccttccatctca 1162
Db 1186 TCTAATAAAACCTTCCATCTCA 1207

RESULT 5
US-09-474-436-8324
; Sequence 8324, Application US/09474436
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
```







QY	600	acgagttgctcgacttaattctctaccttggaaccacccccgacaaagacactgaggcagggtt	659
DB	647	acgagttgctcgacttaattctctaccttggaaccacccccgacaaagacactgaggcagggtt	706
QY	660	caagcgtgcaatggatcttggttcaagagaggtttctgcagagattagatttctaccaaca	719
DB	707	caagcgtgcaatggatcttggttcaagagaggtttctgcagagattagatttctaccaaca	766
QY	720	cagctggctgcagcccgccgtcttggtgaagagagcccttcccagcgattccagtgga	779
DB	767	cagctggctgcagcccgccgtcttggtgaagagagcccttcccagcgattccagtgga	826
QY	780	cccaagtggagagattggtgaaactggcaactggcaatgctatctccctgggaaggagcatctcta	839
DB	827	cccaagtggagagattggtgaaactggcaactggcaatgctatctccctgggaaggagcatctcta	886
QY	840	ccacctggaaatctggcgctgtccacctccagtgccaattcttcttggatctacactgacca	899
DB	887	ccacctggaaatctggcgctgtccacctccagtgccaattcttcttggatctacactgacca	946
QY	900	ggctggacagtgccggaatacagtggtgcccaagagagcccacacatttccaaaagccgct	959
DB	947	ggctggacagtgccggaatacagtggtgcccaagagagcccacacatttccaaaagccgct	1006
QY	960	gccccggcagagccatggcggggtcttcgggacagagccctggaccaggttcagtgggat	1019
DB	1007	gccccggcagagccatggcggggtcttcgggacagagccctggaccaggttcagtgggat	1066
QY	1020	ccctggctgcatcttgctccatgcaagcggttcaattggcggtcacgcgaccccgagaggg	1079
DB	1067	ccctggctgcatcttgctccatgcaagcggttcaattggcggtcacgcgaccccgagaggg	1126
QY	1080	tgcttgagcatggcccggtgccaccttggccagcgctcatcactctccacaaattctcta	1139
DB	1127	tgcttgagcatggcccggtgccaccttggccagcgctcatcactctccacaaattctcta	1186
QY	1140	gtctaataaaaccttccattcta	1162
DB	1187	gtctaataaaaccttccattcta	1209

**RESULT 8**

```

US-09-359-922-3400
; Sequence 3400, Application US/09359922
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
; TITLE OF INVENTION: LIBRARIES
; FILE REFERENCE: 20411-752CON1
; CURRENT APPLICATION NUMBER: US/09/359,922
; CURRENT FILING DATE: 1999-07-22
; EARLIER APPLICATION NUMBER: US 09/205,155
; EARLIER FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 13203
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3400
; LENGTH: 1297
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-359-922-3400

```

Query Match 65.9%; Score 773; DB 50; Length 1297;

Qy 1 gagctgtattggagacaccgtctctcgcggcgctcttaacgctgtgtgtgcgcgcgcca 60  
|||||  
Db 22 gagctgtattggagacaccgtctctcgcggcgctcttaacgctgtgtgtgcgcgcgcca 81  
|||||  
Qy 61 cccctgtataccgggacccgcatagtctcgttcagaggtccgcccccaaaagatcc 120  
|||||

Db	82	ccctgtataccggcacccgcgcatgtctcggtccagagttccgttcccgcccccaaacgatcc	141
Qy	121	cgagcaaacatcatggcacgcgccccgaaatcgggagcgcaaatggcactttccactgcgac	180
Db	142	cgagcaaacatcatggcacgcgccccgaaatcgggagcgcaaatggcactttccactgcgac	201
Qy	181	gaggcactggcatgcgcactgtttcgctctcttcgcgagttaccgggagtcgagagattgtg	240
Db	202	gaggcactggcatgcgcactgtttcgctctcttcgcgagttaccgggagtcgagagattgtg	261
Qy	241	cgaccgccgatccgaaaaatcgcttccgtgtgacatcggttggagcgtggggcgag	300
Db	262	cgaccgccgatccgaaaaatcgcttccgtgtgacatcggttggagcgtggggcgag	321
Qy	301	tacgacctgggagacaccgatatgaccatcacagagtgcttttcacagagaccatgagc	360
Db	322	tacgacctgggagacaccgatatgaccatcacagagtgcttttcacagagaccatgagc	381
Qy	361	tcctgtcccctgggaggcgtggagaccaaagctgagcagtgcgggactcatctatctg	420
Db	382	tcctgtcccctgggaggcgtggagaccaaagctgagcagtgcgggactcatctatctg	441
Qy	421	cacttc-gggcacaaactgctggccagttgctggcactagtadaagagacacatggt	479
Db	442	cacttcggggacaaactgctggccagttgctggcactagtadaagagagacacatggt	501
Qy	480	-gggcacctctatgacaagatgtagaactttgtggagaggtggatgctgtggaca	538
Db	502	ggggcacctctatgacaagatgtagaactttgtggagaggtggatgctgtggaca	561
Qy	539	atgggatctcccagtgggcagagggggagcctcgatatagcactgacaaactacctgagt	598
Db	562	atgggatctcccagtgggcagagggggagcctcgatatagcactgacaaactacctgagt	621
Qy	599	cagcagttgctgacttaactctactctggaacccaccccgacacagacactgagcaggt	658
Db	622	cagcagttgctgacttaactctactctggaacccaccccgacacagacactgagcaggt	681
Qy	659	tcaagcgtgcaatggatctggttccaagaggagtttctgcagagattagatttaccacac	718
Db	682	tcaagcgtgcaatggatctggttccaagaggagtttctgcagagattagatttaccacac	741
Qy	719	acagcttgctgcagccgggcttggtggaagagagcccttgcacagcagattccagttg	778
Db	742	acagcttgctgcagccgggcttggtggaagagagcccttgcacagcagattccagttg	801
Qy	779	acccaagtggagagattgttggaaactggcgaaaggctgcattgcttccttggaaaggacatctct	838
Db	802	acccaagtggagagattgttggaaactggcgaaaggctgcattgcttccttggaaaggacatctct	861
Qy	839	accaactggaaatctgggctgtcccctccaagtggccaatctttttgttatctacactgacc	898
Db	862	accaactggaaatctgggctgtcccctccaagtggccaatctttttgttatctacactgacc	921
Qy	899	aggctggacagtgcgaaatcacagtgtgtccaaaggagcccactcattccaagacgggc	958
Db	922	aggctggacagtgcgaaatcacagtgtgtccaaaggagcccactcattccaagacgggc	981
Qy	959	tgcacctgcagagcagatggcggggtcttctcgggagcgagggcccttgaccaggtctggtggga	1018
Db	982	tgcacctgcagagcagatggcggggtcttctcgggagcgagggcccttgaccaggtctggtggga	1041
Qy	1019	tcactggctgcatctctgtccatgcaagcggcttcaattggcggtca	1064
Db	1042	tcactggctgcatctctgtccatgcaagcggcttcaattggcggtca	1081

RESULT 9  
US-09-359-922-3400  
; Sequence 3400, Application US/09359922A  
; GENERAL INFORMATION:  
; APPLICANT: Leshkowitz, Dena  
; APPLICANT: Liu, Jin









---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 14:44:37 ; Search time 3022.95 Seconds  
(without alignments)  
1465.077 Million cell updates/sec

Title: US-09-215-435-117  
Perfect score: 1173  
Sequence: 1 gagctgcttatggacacg.....tccatctcaaaaaaaaaa 1173

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4538634 seqs, 1887831982 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

EST.\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*

44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_gss1:\*  
80: gb\_gss2:\*  
81: gb\_gss3:\*  
82: gb\_gss4:\*  
83: em\_gss1:\*  
84: em\_gss2:\*  
85: em\_gss3:\*  
86: em\_gss4:\*  
87: gb\_gss5:\*  
88: gb\_gss6:\*  
89: gb\_gss7:\*  
90: gb\_gss8:\*  
91: gb\_gss9:\*  
92: em\_gss5:\*  
93: em\_gss6:\*  
94: em\_gss7:\*  
95: em\_gss8:\*  
96: em\_gss9:\*  
97: em\_gss10:\*  
98: em\_gss11:\*  
99: gb\_gss10:\*  
100: gb\_gss11:\*  
101: em\_gss12:\*  
102: gb\_gss12:\*  
103: gb\_gss13:\*  
104: gb\_gss14:\*  
105: gb\_gss15:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Match	Length	ID	Description
No.					
c 1	555	47.3	682	62	AI911227
					AI911227 wg33b02.x





```

Qy 1157 atctca 1162
Db 15 ATCTCA 10

RESULT 2
LOCUS AI700463 541 bp mRNA EST 03-JUN-1999
DEFINITION wd14c04.x1 Soares_NFL_T-GBC_S1 Homo sapiens cDNA clone
IMAGE:2328102 3' similar to WP:K08H10.8 CE18880 YEAST HYPOTHETICAL
PROTEIN YEY6 LIKE ;, mRNA sequence.
ACCESSION AI700463
VERSION AI700463.1 GI:4988363
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 541)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1998 this sequence version replaced gi:3136831.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 454.
FEATURES
source
1. 541
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2328102"
/clone_lib="Soares_NFL_T-GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL9W, testis NHT, and B-cell
NCI_CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "
110 a 148 c 162 g 121 t

Query Match 45.3%; Score 531; DB 50; Length 541;
Best Local Similarity 100.0%; Pred. No. 2.8e-240;
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 628 aaccaccgccgaagacactgagcgaggttcaagctgcaatgactgtgttcaagag 687
Db 541 AACCAACCCGACCAAGACACTGAGCGAGGTTCAAGCGTGCANAGATCTGTTCAAGAG 482

Qy 688 gagttctgcagagattagattttaccacacacagctggtgccagccggcgcttgggtg 747
Db 481 GAGTTCTGCAAGATTAGATTCTTACCACACACAGCTGCTGCCAGCCGGCGCTGGTG 422

Qy 748 gaacagggccctgccacgattccagtgacccaagtggagagattgtggaactggcg 807
Db 421 GAAGAGGCCCCGTGGCCAGCGATTCCAGGTGACCCCAAGTGGAGAGATTGTGGAATCTGGC 362

Qy 808 aaagtgcatgtccctggaagagagcatctaccacctgggaatctgggctccctcca 867
Db 361 AAAGGTGATCTCCCTGGAAGAGAGCATCTCTACCACCTGGATCTGGGCTGTCCCTCCA 302

```

```

Qy 868 gtggccatctcttttttatctacactgaccagctggacagctggcggaatacagtgtgtg 927
Db 301 GTGGCCATCTCTTTGTTATCTACACTGACCAGGCTGGACAGTGGCGGAATACAGTGTGTG 242

Qy 928 ccaagagagcccaactcattccaaagccgctgcccctgcagagccatggcggtgctt 987
Db 241 CCAAGAGAGCCCCACTCATTCCAAAGCCGGCTGCCCTGCCAGAGCCATGGCGGGTCTT 182

Qy 988 cgggacagagccctggaccaggtcagtgggatccctgctgcctcttcctcctcgaagc 1047
Db 181 CGGACGAGGCCCTGGACCAGGTCACTGGATCCCTGGCTGCATCTTCGTCCTCAAGC 122

Qy 1048 ggcttcattggcggtcacccgacccgagaggtgcttccttgagcatggccctgcacattg 1107
Db 121 GGCCTTCATTGGCGGTCAACCGCACCGAGAGGGTGCCTTGAGCATGGCCGCTGCCACCTTG 62

Qy 1108 gccagagctcatcctccacaaaattctctagcttaataaaaccttccat 1158
Db 61 GCCCAGGCGCTACATACCTCCCAAAATCTCTAGTCTAATAAAACCTTCCAT 11

RESULT 3
AW025001/c
LOCUS AW025001 609 bp mRNA EST 27-OCT-1999
DEFINITION wu70h04.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:2525431 3'
similar to WP:K08H10.8 CE18880 YEAST HYPOTHETICAL PROTEIN YEY6 LIKE
; , mRNA sequence.
ACCESSION AW025001
VERSION AW025001.1 GI:5878531
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 609)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187100.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

```

Seq primer: -40UP from Gibco  
High quality sequence stop: 450.

FEATURES  
source

1. 609  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2525431"  
/clone\_lib="NCI\_CGAP\_Kid3"  
/lab\_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer,  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not  
I and Eco RI sites of the modified pT73 vector. mRNA  
source: 2 pooled kidneys. Library went through one round  
of normalization. Library constructed by Bento Soares and  
M. Fatima Bonaldo. "

BASE COUNT 125 a 166 c 181 g 137 t  
ORIGIN



ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 555)  
AUTHORS Jia,L.B., Young,M.F., Touchman,J.W., Bouffard,G.G.,  
Beckstrom-Sternberg,S.M., Green,E.D., Powell,J.I., Yang,L.M.,  
Robey,P.G., Hotchkiss,R.N. and Francomano,C.A.  
SCAP: The Skeletal Genome Anatomy Project  
Unpublished (1997)  
JOURNAL  
COMMENT On May 9, 1996 this sequence version replaced gi:1132700.  
Contact: Libin Jia  
Medical Genetics Branch  
National Human Genome Research Institute  
10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA  
Tel: 301-402-4877  
Fax: 301-496-7157  
Email: libin@helix.nih.gov  
DNA Sequencing and analyses by National Institutes of Health  
Intramural Sequencing Center (NISC).  
Plate: 18 row: e column: 03  
Seq primer: -21M13 forward primer (ABI).  
Location/Qualifiers  
1. .555  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NHTBC.cn18e03"  
/clone\_lib="Normal Human Trabecular Bone Cells"  
/sex="Female"  
/tissue\_type="Bone"  
/cell\_type="Trabecular Bone Cells"  
/lab\_host="SURE"  
/note="Organ: Hip; Vector: pbluescript; Site\_1: EcoRI"  
BASE COUNT 112 a 147 c 162 g 134 t  
ORIGIN  
Query Match 40.4%; Score 474; DB 51; Length 555;  
Best Local Similarity 100.0%; Pred. No. 2.2e-213;  
Matches 474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 689 agttctgcagagattagatttctaccacacagctgctgccagcccgcccttggtg 748  
Db 492 AGTTTCGACAGATTAGATTCTACCAACACACAGCTGGTCCGAGCCCGGCTTGGTG 433  
QY 749 aagagggcccttgcacagcttccagtgagcccaagtgagagattgtggaactgcga 808  
Db 432 AAGAGGCCCTTGCCACAGATTCCAGGTGGACCCCAAGTGGAGAGATTGTGGAACCTGCGCA 373  
QY 809 aaggtgcattccctggaagagagatctctaccacctggaatctgggctgtccctccag 868  
Db 372 AAGGTGCATTCCCTGGAAGAGAGATCTCTACCACTGGATCTGGGCTGTCCCTCCAG 313  
QY 869 tggccattcttctttgttatctacactgaccaggtgagcagtggtgacagtggtgacagtgtgc 928  
Db 312 TGGCCATCTCTTCTTTGTATCTACACTGACCAGGCTGGACAGTGCGCAATACAGTGTGC 253  
QY 929 ccaaggagccccactatccaaagccggtgcccctgccagaccatgcgctgttc 988  
Db 252 CCAAGGAGCCCCACTCATATCCAAAGCCGGCTGCCCTGCCAGAGCCATGCGGGGTCTC 193  
QY 989 ggagcaggcccttgaccagtgagtgatccctggctgctcatcttcgtccatgcagagc 1048  
Db 192 GGGACGAGGCCCTGGACACAGTCAAGTGGGATCCCTGGCTGCATCTTCGTCCATGCAAGCG 133  
QY 1049 gcttcattggcgttcaccgcaccgcagagaggtgcttgtagcattggccctggtccaccttg 1108  
Db 132 GCTTCATTGGCGGTACCGCACCGAGAGGGTGCCTTGAGCATGGCCCGTGGCCACCTTGG 73  
QY 1109 cccagcgtctacacctcccaaaatctcttagtctataaaaccttcattctca 1162  
Db 72 CCCAGCGCTCATACCTCCCAAAATCTCCTAGTCTAATAAAACCTTCCATCTCA 19

RESULT 6  
AI554935/c  
LOCUS  
DEFINITION te48el2.x1 Soares\_NhMPu\_S1 Homo sapiens CDNA clone IMAGE:2089966  
3' similar to WP:C27H6.5 CE08429 YEAST HYPOTHETICAL PROTEIN YEY6  
LIKE : mRNA sequence.  
ACCESSION AI554935  
VERSION AI554935.1 GI:4487298  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 477)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On May 7, 1998 this sequence version replaced gi:3121644.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 1109 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 441.  
Location/Qualifiers  
1. .477  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2089966"  
/clone\_lib="Soares\_NhMPu\_S1"  
/tissue\_type="Pooled human melanocyte, fetal heart, and  
pregnant uterus"  
/lab\_host="DH10B"  
/note="Organ: mixed (see below); Vector: pT73D-Pac  
(Pharmacia) with a modified polylinker; Site\_1: Not I;  
Site\_2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NbHM, pregnant uterus  
NBHPU, and fetal heart NBHH19W) were mixed, and ss circles  
were made in vitro. Following HAP purification, this DNA  
was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from pools of  
5,000 clones made from the same 3 libraries. The pools  
consisted of I.M.A.G.E. clones 260232-265223,  
340488-345479, and 484488-489479."  
BASE COUNT 101 a 132 c 147 g 97 t  
ORIGIN  
Query Match 40.3%; Score 473; DB 48; Length 477;  
Best Local Similarity 100.0%; Pred. No. 6.6e-213;  
Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 689 agttctgcagagattagatttctaccacacagctggtgccagcccgcccttggtg 748  
Db 473 AGTTTCGACAGATTAGATTCTACCAACACACAGCTGGTCCGAGCCCGGCTTGGTG 414  
QY 749 aagagggcccttgcacagcttccagtggtgacccaagtgagagattgtggaactgcga 808  
Db 413 AAGAGGCCCTTGCCACAGATTCCAGGTGGACCCCAAGTGGAGAGATTGTGGAACCTGCGCA 354  
QY 809 aaggtgcattccctggaaggagagcatctctaccacctggaatctgggctgtccctccag 868  
Db 353 AAGGTGCATTCCCTGGAAGGAGCATCTCTACCACTGGAATCTGGCTGTCCCTCCAG 294  
QY 869 tggccattcttctttgttatctacactgaccaggtggtgacagtggtggaatacagtgtgc 928  
Db 293 TGGCCATCTCTTCTTTGTATCTACACTGACCAGCTGGACAGTGGCGAATACAGTGTGC 234  
QY 929 ccaaggagccccactatccaaagccggtgcccctgccagaccatgcgctgttc 988

us-09-215-435-117.oliq.rst

QY	722	gttggtgccaagccgggcttggtggaagagcccttggcccagcagcattccaggtggacc	781
Db	442	GCTGGCTGCACGCGGGCGCTTGGTGGAGAGAGCCCTTGCACAGCGATTCACAGGTGGACC	383
QY	782	caagtgcagagatttggaactggcgaaaggtgcattgctcctgggaagagcattctctacc	841
Db	382	CAAGTGGAGAGATTGTGGAACTGGCGAAAGGTGCATGTCCCTGGAAGGACATCTCTACC	323
QY	842	acctggaattctgggtgctccctccagtgggccattcttcttattctacacgtgaccagg	901
Db	322	ACCTGGAATCTGGGCTGTCCCTCCAGTGGCCATCTCTTTCTTATCTACATGACACAGG	263
QY	902	ctggacagtggcgaaatcacagtgtgtcccaagagagcccaactcattccaaagccgcgtgc	961
Db	262	CTGGACAGTGGCGAAATACAGTGTGTGCCCAAGGAGCCCCACATTCATTCCAAAGCCGGCTGC	203
QY	962	ccctgcagagccatggcggggtcttcgggagagagccctggaccaggtcagtgggatcc	1021
Db	202	CCTGCCAGAGCCATGGCGGGGTCTTCGGGACGAGGCCCTGGACCAAGTCACTGGGATCC	143
QY	1022	ctggctgcattcttcgtccatgcgaagcgggttcaattggcgggtcacccagcagagaggtg	1081
Db	142	CTGGCTGCATCTTCGTCCATGCAGCGGGTTCATTGTCGGGTCAACCACCCGAGAGGGTG	83
QY	1082	ccttgagcatggccgtgcacattggccagcgcctacacctccacaaatctctcagt	1141
Db	82	CCTTGACATGGCCGGCCGGCCACCTTGGCCCGAGCGCTCATACCTCCACAAATCTCCTAGT	23
QY	1142	ctaataaaacctctccattca	1162
Db	22	CTAATAAAACCTTCCATCTCA	2
RESULT 8			
AI277497/c			
LOCUS			
DEFINITION			
gi96008.x1 Soares_NhhMPu_S1 Homo sapiens cDNA clone IMAGE:1880175			
3', similar to WP:C27H6.5 CE08429 YEAST HYPOTHETICAL PROTEIN Y5Y6			
LIKE : mRNA sequence.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			

NbHPV, and fetal heart NbH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 90 a 122 c 136 g 87 t  
ORIGIN

Query Match 34.0%; Score 399; DB 44; Length 435;  
Best Local Similarity 100.08; Pred. No. 5.6e-178; Indels 0; Gaps 0;  
Matches 399; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 764 agcattccagtgagacacaaagtggagagattgtgaaactgacggaagtgcatgcccct 823  
|||||  
Db 401 AGCGATTCCAGGTGACCCCAAGTGGAGAGATTGTGAACTGGCGAAAGGTGATGCCCT 342  
|||||  
Qy 824 gaaaggagcatctacacacatgggaatctgggtgtccctccatggccatctcttgg 883  
|||||  
Db 341 GGAAGGAGCATCTACCACTGGAACTGGGCTGTCCCTCCAGTGGCCATCTCTTTG 282  
|||||  
Qy 884 ttatctacatgacacagctgacagtgagcgaatacagtggtgtcccaaggagcccaact 943  
|||||  
Db 281 TTATCTACACTGACCAAGCTGGACAGTGGCGAATACAGTGTGTGCCCAAGAGCCCCACT 222  
|||||  
Qy 944 cattccaagccggctgcccctgcagagccatggcggtgttcgggacgagccctgg 1003  
|||||  
Db 221 CATTCCAAAGCGGCTGCCCTGCCAGAGCCATGGCGGGTCTTCGGACGAGGCCCTGG 162  
|||||  
Qy 1004 accaggtcagtggtgattccctggctgcatcttctgctccatgcaagcggcttcattggcggtc 1063  
|||||  
Db 161 ACCAGGTGAGTGGGATCCCTGGCTGCATCTTCGTCATGCAAGCGCTTCATTGGCGGTC 102  
|||||  
Qy 1064 accgcacccgagaggtgcttgacatggcccggtgacacattggccagcgcctatacc 1123  
|||||  
Db 101 ACCGACCCGAGAGGGTGCTTGACATGGCCCCGGTGGCCACTTGGCCAGCGCTCATACC 42  
|||||  
Qy 1124 tcccacaatctcctagtgtaataaaaccttccatctca 1162  
|||||  
Db 41 TCCCAAAATCTCCTAGTCTAATAAAACCTTCCATCTCA 3  
|||||

RESULT 9  
AI936652/c  
LOCUS  
DEFINITION wp58b08.x1 NCI-CCAP\_Brn25 Homo sapiens cDNA clone IMAGE:2466903 3', similar to WP:K08H10.8 CE18880 YEAST HYPOTHETICAL PROTEIN YEF6 LIKE  
; mRNA sequence.  
ACCESSION AI936652  
VERSION AI936652.1 GI:5675533  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 536)  
AUTHORS NCI/NINDS-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index  
JOURNAL Unpublished (1998)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187046.  
Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40Up from Gibco  
High quality sequence stop: 480.  
Location/Qualifiers

FEATURES  
source

1. .536  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2466903"  
/clone\_lib="NCI-CCAP\_Brn25"  
/tissue\_type="anaplastic oligodendroglioma"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGCGGCGCATAGGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 110 a 147 c 164 g 114 t 1 others  
ORIGIN

Query Match 32.4%; Score 380; DB 63; Length 536;  
Best Local Similarity 99.68; Pred. No. 5.2e-169;  
Matches 480; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 639 ccaagacactgagcgaggttcaagcgtgcaatggatctggttcaaggaggtttctgca 698  
|||||  
Db 525 CCAAGACACTGAGCGAGGGTTCAAGCGTGCATGTGTTCAAGAGAGTTTCTGCA 466  
|||||  
Qy 699 gagattagattctacacacacagctggtccagccggccttggtggaagagccct 758  
|||||  
Db 465 GAGATTAGATTCTACCAACACAGCTGGCTGCCAGCCGGCCTTGTGGAGAGGCCCT 406  
|||||  
Qy 759 tgcccagcgtattccaggtggacccaagtggagagattgtggaactggcgaaggtgcatg 818  
|||||  
Db 405 TGCCGAGCGATTCCAGGTGGACCCCAAGTGGAGAGATTGTGGAAGTGCATG 346  
|||||  
Qy 819 tccctggaagagagcatctctaccacctggaatctgggtgtccctccagtgccattt 878  
|||||  
Db 345 TCCCTGGAAGAGCATCTCTACCACTGGAATCTGGCTGTCCCCCTCCAGTGGCCATCTT 286  
|||||  
Qy 879 cttgtattctacactgacaggtggacagtggaatggaatacagtggtgccaagagacc 938  
|||||  
Db 285 CTNCTTTATCTACATGACAGGCTGGACAGTGGCAATACAGTGTGTGCCAAGAGGCC 226  
|||||  
Qy 939 ccaactcattccaaagccggtgcccctgccagagcagcatggcggtctctcgggagagcc 998  
|||||  
Db 225 CCATCATTCCAAAGCCGGCTGCCCTGCCAGAGCCATGGCGGGGTCTTCGGGAGAGGC 166  
|||||  
Qy 999 cctggaacaggttcagtggtggtatccctgggtggtggttcttccatgcaagcgttccattgg 1058  
|||||  
Db 165 COTGACCAGGTTCAGTGGGATCCCTGGCTGCATCTTCGTCCATGCAAGCGGCTTCATTGG 106  
|||||  
Qy 1059 caggtccacccagcagagaggtgcttgagcatggccgtgcccacttgccacagcgtc 1118  
|||||  
Db 105 CGGTCAACCGACCCGAGAGGGTGCCTTGAGCATGGCCCGCTGACCTTGGCCCGCGGTC 46  
|||||

Qy 1119 at 1120  
||  
Db 45 AT 44

RESULT 10  
AI095901/c  
LOCUS  
DEFINITION AI095901 494 bp mRNA EST 05-OCT-1998  
qB21b03.x1 Soares\_pregnant\_uterus\_NbHPV Homo sapiens cDNA clone  
IMAGE:1696877 3' similar to WP:C27H6.5 CE08429 YEAST HYPOTHETICAL

PROTEIN YEY6 LIKE ; mRNA sequence.

AI095901  
AI095901.1 GI:3434877

EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 494)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
On Jan 9, 1998 this sequence version replaced gi:936710.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1161 Std Error: 0.00  
Seq primer: -40ml3 fwd. Et from Amersham  
High quality sequence stop: 370.  
Location/Qualifiers  
1. .494  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1696877"  
/clone\_lib="Soares\_pregnant\_uterus\_NbHPU"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pT7T3-Pac; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGGAAGATCGGGCGGCGCTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."

BASE COUNT 102 a 131 c 147 g 114 t

ORIGIN  
Query Match 32.2%; Score 378; DB 42; Length 494;  
Best Local Similarity 99.6%; Pred. No. 4.6e-168;  
Matches 478; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 683 aagagggtttctcagagattagattttctaccacacagctggctccagccggccct 742  
DB 494 AAGAGGAGTTTCTGCAGAGATTAGATTCTACCAACACAGCTGGCTGCCAGCCGGCCCT 435  
QY 743 tggtagaagagccctgtccagcgattccagggtggaccacagtgagagattgtggaac 802  
DB 434 TGGTGGAGAGGCCCTTGGCCAGCGATTCCAGGTGGACCAAGTGAGAGATTGTGGAAC 375  
QY 803 tggcgaaggtgcatgtccctgggaagagcatctctaccacctggaatctgggctgcc 862  
DB 374 TGGCAAAAGGTGCATGTCCCTGGAAGGAGCATCTCTACCACCTGGAAATCTGGGCTGCC 315  
QY 863 ctccagtgcccatctctttttatctacactgaccaggtggacagtggcgcaatacagt 922  
DB 314 CTCCAAATGGCCATCTCTTTGTATCTACACTGACAGGCTGGAGCTGGCGAATACAGT 255  
QY 923 gtgtgccaaagagagcccaactcattccaaagccggtgcccctgcagagccatggcggg 982  
DB 254 GTGTGCCCAAGAGAGCCCACTTCATCCAAAGCCGGCTGCCCTGCCAGAGCATGGGGG 195  
QY 983 gtctcggagagagcccttgaccaggtcagtgggatccctggctgcatcttcctcatg 1042  
DB 194 GTCTTCGGGACAAAGGCCCTTGACACAGGTCACTGGGATCCCTGGCTGCATCTTCGTCCATG 135  
QY 1043 caagcggtcttattggcggtaccgcacccagagagggtgcttggagcatggccgtgcca 1102

Db 134 CAACGGCTTCATTGGCGGTACCGCACCGAGAGGGTGCCTTGAGCATGGCCGTGCCA 75  
QY 1103 ccttgccccagcgctcattaccctcccaaaatctcctagtcttaataaaaccttcacatca 1162  
Db 74 CTTTGGCCAGCGCTCATACCTCCCAAAATCTCCTAGTCTAATAAAACCTTCCATCTCA 15

RESULT 11  
A1138745/c  
A1138745 LOCUS  
DEFINITION q98nc10.x1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:1737522  
3' similar to WP:C27H6.5 CE08429 YEAST HYPOTHETICAL PROTEIN YEY6  
LIKE ; mRNA sequence.  
ACCESSION A1138745  
VERSION A1138745.1 GI:3644717  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 492)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced gi:2285425.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www.bld.llnl.gov/bbrp/image/image.html

Insert Length: 1068 Std Error: 0.00  
Seq primer: -40ml3 fwd. Et from Amersham  
High quality sequence stop: 444.  
Location/Qualifiers  
1. 492  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1737522"  
/clone\_lib="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGGCCCAATTTTTTTTTTTT 3'], Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 101 a 133 c 150 g 108 t

ORIGIN  
Query Match 32.1%; Score 376; DB 42; Length 492;  
Best Local Similarity 100.0%; Pred. No. 4e-167;  
Matches 376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 685 gaggagtcttcagagattagattttctaccacacagctggctgccagccggcgcttg 744  
DB 492 GAGGAGTTTCTGCAGAGATTAGATTCTACCAACACAGCTGGCTGCCAGCCGGCGCTTG 433  
QY 745 gtggaagagggcccttgcacagcgattccaggtggaccacagtgagagattgtggaactg 804



/note="Organ: brain; Vector: PCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 1.33 kb. Tumor types include: meningioma, oligodendroglioma, astrocytoma (grade II), medulloblastoma, astrocytoma (grade IV). Life Technologies catalog #: 11544-012"

BASE COUNT 87 a 115 c 132 g 95 t 2 others  
ORIGIN

Query Match 29.2%; Score 343; DB 43; Length 431;  
Best Local Similarity 99.7%; Pred. No. 1.5e-151;  
Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 779 acccaagtgaagatgtggaactggcgaagtgcatgtccctgaaagagcatctct 838  
|||||  
Db 394 ACCCAAGTGGAGAGATTGTGGAAGTGGCGAAGGTGATGTCCTCGAAGAGCATCTCT 335  
|||||

QY 839 accaactggaatctgggtgtccctccctcagtgccatctctttttgttatctacactgacc 898  
|||||  
Db 334 ACCACCTGGAATCTGGGCTGTCCCTCCAGTGGCCATCTTCTTGTATCTACACTGACC 275  
|||||

QY 899 aggtgacagtggcgaatacagtggtgcccgaaggagcccaactcattccaaagccggc 958  
|||||  
Db 274 AGGTGGACAGTGGCGAATACAGTGTGTGCCAAGGAGGCCCACTCATTCCTCAAGCCGGC 215  
|||||

QY 959 tgcccttgccagagccatggcgggtcttcgggacgagggccctggaccaggtcagtgagg 1018  
|||||  
Db 214 TGCCCTTGCCAGAGCCATGGCGGGTCTTCGGGACGAGGCCCTGGACAGTCACTGCGA 155  
|||||

QY 1019 tccctggctgcatcttcgtccatcgaagcggttcattgctgagccacccagccagagg 1078  
|||||  
Db 154 TCCCTGGCTGCATCTTCGTCCATCAAGCGGCTTCATTGGCGGTACCGCACCCGAGAGG 95  
|||||

QY 1079 gtgcttgagatggccgtgccccttggccagcgtcatcattcccaaatctcct 1138  
|||||  
Db 94 GTGCTTGAGCATGGCGGCGGACCTTGCGCCAGCGCTATACCTCCCAAAATCTCCT 35  
|||||

QY 1139 agtctaataaaccttccatctcaaaaaa 1172  
|||||  
Db 34 AGTCTAATAAACCTTCATCTCAAAAAA 1  
|||||

RESULT 14  
AI206725/c  
LOCUS  
DEFINITION  
qf61c07.x1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:1754508  
3', similar to WP:C27H6.5 CE08429 YEAST HYPOTHETICAL PROTEIN Y66  
LIKE ; mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 428)  
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Seq primer: -400p from Gibco

High quality sequence stop: 318.  
Location/Qualifiers  
1..428

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone IMAGE:1754508"  
/clone\_lib="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from Clontech  
Laboratories, Inc., and primed with a Not I - oligo(dT)  
primer [5'  
TGTTCACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTTTT 3'].  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization to Cot5, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 90 a 116 c 132 g 89 t 1 others  
ORIGIN

Query Match 29.2%; Score 342; DB 43; Length 428;  
Best Local Similarity 100.0%; Pred. No. 4.5e-151;  
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 821 cctgggaagagcatctctaccacctggaatctgggtgtccctccagtgccatctct 880  
|||||  
Db 350 CCTGGGAAGAGCATCTCTACCACCTGGATCTGGGCTGTCCCTCCAGTGGCATCTCT 291  
|||||

QY 881 ttgttatctacactgaccaggtggacagtggcgaatacagtggtgcccaggagagccc 940  
|||||  
Db 290 TTGTTATCTACACTGACAGGCTGGACAGTGGCGAATACAGTGTGTGCCAAGAGCCCC 231  
|||||

QY 941 actcattccaaagccggtgcccctccagagccatgggggtcttcgggacgagccc 1000  
|||||  
Db 230 ACTCATTCCAAAGCCGGCTGCCCTGCCAGAGCCATGGCGGGGTCTTCGGGACGAGCCCC 171  
|||||

QY 1001 tggaccaggtcagtggtggtccctgggtgcatcttcctcattgcaagcggctcattggcg 1060  
|||||  
Db 170 TGGACCAAGTCACTGGGATCCCTGGCTGCATCTTCCTCCATGCAAGCGGCTTCATTGGCG 111  
|||||

QY 1061 gtccagcaccagagaggtgcttgagatggcccgctgcccacttgccacgctgcccagcgtcat 1120  
|||||  
Db 110 GTCACGCGACCCGAGAGGCTGCTTGAGCATGGCCGCTGCCACCTTGGCCCGACGCTCAT 51  
|||||

QY 1121 acctcccaaatctcctagtctaataaaaccttccatctca 1162  
|||||  
Db 50 ACCTCCCAAAATCTCCTAGTCTATATAAACCTTCCATCTCA 9  
|||||

RESULT 15  
AA972697/c  
LOCUS  
DEFINITION  
Op91d03.s1 Soares\_NFL\_T\_GBC.S1 Homo sapiens cDNA clone  
IMAGE:1584197 3', similar to WP:C27H6.5 CE08429 YEAST HYPOTHETICAL  
PROTEIN Y66 LIKE ; mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 445)  
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Feb 13, 1998 this sequence version replaced gi:2619221.  
Contact: Robert Strausberg, Ph.D.



Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 761 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 412.  
FEATURES  
Location/Qualifiers  
1..445  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1584197"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NbHL19W, testis NHT, and B-cell  
NCLCGAP\_GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo. "  
BASE COUNT 91 a 122 c 142 g 90 t  
ORIGIN

Query Match 28.7%; Score 337; DB 40; Length 445;  
Best Local Similarity 100.0%; Pred. No. 1e-148;  
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 730 ccagcccgccgcttggtgaagagcccttgcacagcattccagtgaggcccaagtga 789  
DB 445 CCAGCCCGGCCCTTGTTGAAGAGGCCCTTGCCACGATTCAGGTGGACCCCAAGTGA 386  
QY 790 gagattgtggaactgcgaaagtgcatgtccctggaagagcatctctaccacctgga 849  
DB 385 GAGATTGTGGAAGTGGCAAGAGTGCATGTCCCTGGAAGAGCATCTTACCACCTGGAA 326  
QY 850 tctgggcttccctccagtgccattctcttcttattatatactgaccagcgtgacag 909  
DB 325 TCTGGCTGTCCCTCCAGTGGCCATCTTTTGTATTCTACACTGACCAAGGCTGGACAG 266  
QY 910 tggcgaatacagtggtgcccaagagcccaactcattccaaagccgctgccccctgcca 969  
DB 265 TGGCGAATACAGTGTGTGCCCAAGAGGCCCACTATTCCAAAGCCGCTGCCCTGCCA 206  
QY 970 gagccatgcggggtcttcgggacgagccctggaccagtcagtgggatccctggctgc 1029  
DB 205 GAGCCATGCGGGGTCTTCGGGACGAGGCCCTGGACAGTCAAGTGGGATCCCTGGCTGC 146  
QY 1030 atcttgctccatcaagcggcttcattgcggtcacc 1066  
DB 145 ATCTTCGTCATGCAAGCGGCTTCATTGGCGGTCAAC 109

Search completed: May 1, 2000, 14:44:45  
Job time: 14005 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:42:42 ; Search time 4425.31 Seconds  
(without alignments)  
-538.620 Million cell updates/sec

Title: US-09-215-435-118  
Perfect score: 785  
Sequence: 1 cggaaatccgggagtcgggtg.....tatggaaataaagtttttc 785

Scoring table: OLIGO\_NVC  
Gapop 60.0 , Gapext 60.0  
Searched: 821193 seqs, -1518192014 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pi1.\*
- 8: gb\_pi2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_v1.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pi.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_v1.\*
- 32: gb\_hcg1.\*
- 33: gb\_hcg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_ba1.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_hcg3.\*
- 42: gb\_hcg4.\*
- 43: gb\_hcg5.\*

- 44: gb\_hcg6.\*
- 45: gb\_hcg7.\*
- 46: em\_hcg1.\*
- 47: em\_hcg2.\*
- 48: em\_hcg3.\*
- 49: em\_hum5.\*
- 50: gb\_pi3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	393	50.1	908	40	AF177398	Homo sapi
2	392	49.9	43292	42	AC010643	Homo sapi
C 3	81	10.3	9966	41	AC010524	Homo sapi
4	49	6.2	120400	33	HSJ854E16	Homo sapi
C 5	23	2.9	2053	4	AB008374	Oncorhync
C 6	22	2.8	100490	33	AC007776	Homo sapi
7	20	2.5	376	10	AB024091	Pan trogl
8	20	2.5	376	10	AB024112	Saimiri
9	20	2.5	10782	2	AE001002	Archaeogl
10	20	2.5	22166	43	AC014435	Drosophil
11	20	2.5	34871	43	AC011166	Homo sapi
12	20	2.5	92509	41	AC010844	Drosophil
13	20	2.5	114996	44	AC016738	Homo sapi
14	20	2.5	125032	40	AC007099	Homo sapi
15	20	2.5	160848	43	AC009647	Homo sapi
16	20	2.5	165370	43	AC010843	Drosophil
C 17	20	2.5	225636	44	AC010705	Drosophil
C 18	20	2.5	236763	43	AC010995	Drosophil
19	19	2.4	828	12	AF177399	Mus muscu
20	19	2.4	1258	40	AF096785	Homo sapi
21	19	2.4	2278	9	HUMMHC08	Human cardi
22	19	2.4	2366	9	HSMHGAG1	Human alpha
23	19	2.4	2817	34	AF015539	Mytilus e
24	19	2.4	2984	12	AB006138	Rattus no
25	19	2.4	3508	35	AF053538	Alvinella
26	19	2.4	3870	12	AB004831	Rattus no
27	19	2.4	6345	10	PHU38463	Papio hamad
28	19	2.4	31462	9	HSCAMHCA	Homo sapien
29	19	2.4	33804	11	HS407A10	Human DNA s
30	19	2.4	38041	11	AC002985	Human DNA
31	19	2.4	38596	40	AC004177	Homo sapi
C 32	19	2.4	38786	4	AF170972	Agelaius
33	19	2.4	39978	11	AC005197	Homo sapi
C 34	19	2.4	41069	33	AC004179	Homo sapi
C 35	19	2.4	74596	42	AC009100	Homo sapi
36	19	2.4	84124	32	HSJ9614	Homo sapi
37	19	2.4	86162	11	HSJ988G17	Human DNA
38	19	2.4	91640	11	HS426F10	Human DNA
39	19	2.4	131155	40	AC006137	Homo sapi
40	19	2.4	134292	11	HS271M21	Human DNA
C 41	19	2.4	166686	32	CNS01DTR	Human DNA
42	19	2.4	175625	33	AC007879	Homo sapi
C 43	19	2.4	196287	10	CNS0000B	Human chr
44	19	2.4	215737	32	HSC196A12	Homo sapi
45	19	2.4	279400	33	AC007737	Homo sapi

ALIGNMENTS

RESULT 1  
AF177398  
LOCUS AF177398 908 bp mRNA  
DEFINITION Homo sapiens soggy-1 protein (SGY-1) mRNA, complete cds.  
ACCESSION AF177398  
VERSION AF177398.1 GI:6049611  
PRI 16-OCT-1999





```

phage etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 0021 Length: 1386bp
Contig_ID: 0027 Length: 30604bp
Contig_ID: 0033 Length: 2392bp
Contig_ID: 0040 Length: 7384bp
Contig_ID: 0043 Length: 19908bp
Contig_ID: 0045 Length: 4583bp
Contig_ID: 0049 Length: 1666bp
Contig_ID: 0074 Length: 2312bp
Contig_ID: 0050 Length: 2246bp
Contig_ID: 0052 Length: 10165bp
Contig_ID: 0059 Length: 1218bp
Contig_ID: 0071 Length: 1618bp
Contig_ID: 0079 Length: 7772bp
Contig_ID: 0092 Length: 1978bp
Contig_ID: 0089 Length: 11290bp
Contig_ID: 0091 Length: 1878bp
Contig_ID: 0103 Length: 1878bp
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES             Location/Qualifiers
     source            1..120400
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /chromosome="20"
                       /clone="RP5-854E16"
                       /clone_lib="RPCI-5"
BASE COUNT           32364 c 19430 g 33447 t 12003 others
ORIGIN

Query Match          6.2%; Score 49; DB 33; Length 120400;
Best Local Similarity 100.0%; Pred. No. 7.1e-17;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 acaggctccagagctactcaaggcttcagccgactttctgaaag 255
|||||
DB 3167 ACAGGCTCCAGAGCTACTCAAGGCTTCAGCGACTTTCTCGAAG 3215

RESULT 5
AB008374/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

/codon_start=3
/product="alpha 3 type I collagen"
/protein_id="BAA33381.1"
/db_xref="GI:3641659"
/translation="PGPTGGSGSGERCEHAGCAPGAGPPGADGPGNKGTEGNNG
PKGEAGAPGPGVAGPGQPGAGNSGTGRTAGPAGPPGASGAKGNDGPMGAPGTP
PGSAGPPGAGKEGQGGRTGAGRGEAGAPGPGSGASGAKGNDGPMGAPGTP
GPGIAGQGVGVGGPGGPGGTAGIPTAGPGKQGGVGGSGSGSGSGSGSGSG
APGEAGREGSTGHDGVSQDGGPPGPKGDRGNGNAGSPAGTGPAGPSGASGTGS
RGESGAPGAGNGPSGPPGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG
GSPGPPGAGNGPSGPPGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG
SPGLPGPPGAGGPHVAPINQERKPPGMRGGGGYHRADEPDMNGRDMDVDTLK
SLSKVENIRSPGSKNPKRCMDIRKHKVMSQDGGVQFQSGSGSGSGSGSG
TGCTCVPSQDTPVPMKNYTKNIRKHKVMSQDGGVQFQSGSGSGSGSGSG
TFMRLMSNEASONVTHCKNSIAYMDEATGNLKALLQGAENEIRAEGRSRTYS
SEDGCTSHGTGWTVIDYKTSKTSRLPIIDAPMDVAGPQDFGVEGVPVCFV"
BASE COUNT          420 a 606 c 656 g 371 t
ORIGIN

Query Match          2.9%; Score 23; DB 4; Length 2053;
Best Local Similarity 100.0%; Pred. No. 0.053;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 aggaagaccagagcagcagctg 356
|||||
DB 1184 AGGAGACACAGGAGCAGCAGCTG 1162

RESULT 6
AC007776/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* as soon as it is available and the accession number will
* be preserved. 2868: contig of 2868 bp in length
* 1 gap of unknown length
* 2869 contig of 3981 bp in length
* gap of unknown length
* 6850 11677: contig of 4828 bp in length
* gap of unknown length
* 11678 16813: contig of 5136 bp in length
* gap of unknown length
* 16814 23727: contig of 6914 bp in length
* gap of unknown length
* 23728 37142: contig of 13415 bp in length
* gap of unknown length
* 37143 49015: contig of 11873 bp in length
* gap of unknown length
* 49016 62202: contig of 13187 bp in length
* gap of unknown length
* 62203 79211: contig of 17009 bp in length
* gap of unknown length
* 79212 100490: contig of 21279 bp in length.
* Location/Qualifiers
* 1..100490
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /clone="hCIT.413_A.4"
* /clone_lib="Research Genetics/Cal Tech CITB-HSP-C (plates
195-384)"
* /map="17"
* /chromosome="17"
BASE COUNT 24217 a 25467 c 26408 g 23679 t 719 others
ORIGIN

Query Match 2.8%; Score 22; DB 33; Length 100490;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 565 gctgcaggccatccggatgg 586
|||||
DB 54989 GCGTCAGGCCATCCGGATGG 54968

RESULT 7
AB024091 376 bp DNA PRI 25-SEP-1999
LOCUS Pan troglodytes MHC class II DRB gene (allele:PatrDRB5*01), intron
DEFINITION
ACCESSION AB024091.1 GI:4519947
VERSION
KEYWORDS Pan troglodytes DNA.
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Pan.
REFERENCE
1 (sites)
AUTHORS Kupfermann,H., Satta,Y., Takahata,N., Tichy,H. and Klein,J.
TITLE Evolution of Mhc-DRB introns: implications for the origin of
primates
JOURNAL J. Mol. Evol. 48 (6), 663-674 (1999)
MEDLINE 99246377
REFERENCE
2 (bases 1 to 376)
AUTHORS Kupfermann,H., Satta,Y., Takahata,N., Tichy,H. and Klein,J.
TITLE Direct Submission
JOURNAL Submitted (26-FEB-1999) to the DDBJ/EMBL/GenBank databases. Yoko
Satta, The Graduate University for Advanced Studies; Hayama,
Kanagawa 240-0193, Japan (E-mail:satta@soken.ac.jp,
Tel:81-468-58-1549, Fax:81-468-58-1542)
FEATURES
source
1..376
/organism="Pan troglodytes"
/db_xref="taxon:9598"
gene
1..376
intron
1..376
BASE COUNT 94 a 73 c 95 g 114 t
ORIGIN

Query Match 2.5%; Score 20; DB 10; Length 376;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 ctgctctctctacccgtggt 145
|||||
DB 340 CTGCTCTCTCTACCCGTG 359

RESULT 9
AB001002 10782 bp DNA BCT 15-DEC-1997
LOCUS Archaeoglobus fulgidus section 105 of 172 of the complete genome.
DEFINITION
ACCESSION AE001002 AE000782
VERSION
KEYWORDS Archaeoglobus fulgidus.
SOURCE

```

```

intron
/gene="DRB"
1..376
/gene="DRB"
/note="allele:PatrDRB5*01"
/number=4
BASE COUNT 94 a 75 c 97 g 110 t
ORIGIN

Query Match 2.5%; Score 20; DB 10; Length 376;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 ctgctctctctacccgtggt 145
|||||
DB 340 CTGCTCTCTCTACCCGTG 359

RESULT 8
AB024112 376 bp DNA PRI 25-SEP-1999
LOCUS Saimiri sciureus MHC class II DRB gene (Sasc561), intron 4.
DEFINITION
ACCESSION AB024112
VERSION AB024112.1 GI:4519968
KEYWORDS Saimiri sciureus DNA.
ORGANISM
Saimiri sciureus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
Eutheria; Primates; Platyrrhini; Cebidae; Saimiri.
REFERENCE
1 (sites)
AUTHORS Kupfermann,H., Satta,Y., Takahata,N., Tichy,H. and Klein,J.
TITLE Evolution of Mhc-DRB introns: implications for the origin of
primates
JOURNAL J. Mol. Evol. 48 (6), 663-674 (1999)
MEDLINE 99246377
REFERENCE
2 (bases 1 to 376)
AUTHORS Kupfermann,H., Satta,Y., Takahata,N., Tichy,H. and Klein,J.
TITLE Direct Submission
JOURNAL Submitted (26-FEB-1999) to the DDBJ/EMBL/GenBank databases. Yoko
Satta, The Graduate University for Advanced Studies; Hayama,
Kanagawa 240-0193, Japan (E-mail:satta@soken.ac.jp,
Tel:81-468-58-1549, Fax:81-468-58-1542)
FEATURES
source
1..376
/organism="Saimiri sciureus"
/db_xref="taxon:9521"
gene
1..376
intron
1..376
BASE COUNT 94 a 73 c 95 g 114 t
ORIGIN

Query Match 2.5%; Score 20; DB 10; Length 376;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 ctgctctctctacccgtggt 145
|||||
DB 340 CTGCTCTCTCTACCCGTG 359

RESULT 9
AB001002 10782 bp DNA BCT 15-DEC-1997
LOCUS Archaeoglobus fulgidus section 105 of 172 of the complete genome.
DEFINITION
ACCESSION AE001002 AE000782
VERSION
KEYWORDS Archaeoglobus fulgidus.
SOURCE

```

ORGANISM	Archaeoglobus fulgidus	gene	complement(2144.. .2317)
	Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae; Archaeoglobus.	CDS	/gene-"AF1456"
REFERENCE	1 (bases 1 to 10782)		complement(2144.. .2317)
AUTHORS	Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L., Uterback, T., Cotton, M.D., Spriggs, T., Artiaich, P., Kaine, B.P., Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C., Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O., Woese, C.R. and Venter, J.C.		/note-"hypothetical protein; identified by GeneMark; putative"
TITLE	The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus	trNA	/transl_table=11
JOURNAL	Nature 390 (6638), 364-370 (1997)		/product-"A. fulgidus predicted coding region AF1456"
REFERENCE	98049343		/protein_id-"AAB89795.1"
AUTHORS	2 (bases 1 to 10782)		/db_xref="GI:2649116"
	Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L., Uterback, T., Cotton, M.D., Spriggs, T., Artiaich, P., Kaine, B.P., Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C., Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O., Woese, C.R. and Venter, J.C.		/translation-"MKEKMDVNQWESIRKIIIVRVREAKRLLLETRKLRMEMKSS KIAEMIREDRDAR"
			complement(2652.. .2723)
			/gene-"tRNA-Cys-1"
			/product-"Cys GCA"
			complement(2652.. .2723)
			/gene-"tRNA-Cys-1"
			complement(2811.. .3128)
			/gene-"AF1457"
			complement(2811.. .3128)
			/gene-"AF1457"
			/note-"hypothetical protein; identified by GeneMark; putative"
			/codon_start=1
			/transl_table=11
			/product-"A. fulgidus predicted coding region AF1457"
			/protein_id-"AAB89792.1"
			/db_xref="GI:2649113"
			/translation-"MDARDIVLSVFSAAALVVRWLSLYDRVDMTVFFATLLIAS IITLLISIELRMQRIMDEFKSVKRTIAVNSDDLEGRIERLFIEKVRDLEDKLESTERR MVR"
TITLE	Submitted (15-DEC-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA	gene	3219.. .5444
JOURNAL			/gene-"AF1458"
REMARK	In order to show the genes in ascending order on the genome, the origin of this version has been moved by TIGR to position 2093570 of the original version and the opposite strand is shown from the original version.	CDS	3219.. .5444
COMMENT	On Dec 16, 1997 this sequence version replaced gi:2649104.		/note-"similar to GB:L77117 PID:1592139 percent identity: 48.13; identified by sequence similarity; putative"
FEATURES	Location/Qualifiers		/codon_start=1
source	1.. 10782		/transl_table=11
	/organism="Archaeoglobus fulgidus"		/product-"Arp-dependent RNA helicase, putative"
	/db_xref="taxon:2234"		/protein_id-"AAB89786.1"
gene	complement(159.. .1019)		/db_xref="GI:2649107"
CDS	complement(159.. .1019)		/translation-"MEYVTHPLIKENTIERMYOISIAATATKNTLVVPTGLGKTT TAAIVIASRLNEDGKVLFAPTKPLVQSHARFLKRVLKVEIVSLSGVPEPKRKL WEKARIYVSTPOVENDLLAGRISLEDVILVFDHARAVGNVAVYFAKSLRTAKK PULIATSPGSDPERIMEVIONIGIEALEVRETDWSDVAFYVGGKRLENKYDPEE MKEVKERLQICIKIRKRLRELWIEVPENSSKRDLLALQELAQAAASSQSEIFPEAL SILAETMKLQHAVELIETQGVKAVSYLRKLVREATSGGSKAAKSIYGDPIFFKAVI ALSCKVEHPKLEKLEIKQEFKNPDSRVIVFTNYRDSAEMLYNELSPFPVAKFY GOASRNDKGMKQKEQIETIDKFRGVYKLVATSVGEGDLDPSTDVLFVYEAQVQ IRATQKGTGRTGREGRIIVLVTKCTRDENVYSSMKERKWDKILKRTIDRKQR SIGDVLPETGKIVIVDSRELRSRVKHLREIGAKIRNLEADVADVVSQVAVYERK TVEDFLNSIQKRLFSQVARKLSAYSRPVIIEGENFIRGGVHPNVRGAIASLIID FGIVLRSNARETAELIFAMARQEEKRGVHEHTAKTKRLDKEQEYIVSAISNV GNVIAIIRLDLYFTTIENATIDEELAKVPKVGKKIARRVMTTPTSEAGFYDSES F"
			5471.. .5543
			/gene-"tRNA-Thr-1"
			/product-"Thr GGT"
			5471.. .5543
			/gene-"tRNA-Thr-1"
			complement(5605.. .5877)
			/gene-"AF1459"
			complement(5605.. .5877)
			/gene-"AF1459"
			/note-"similar to GB:L77117 SP:Q57735 PID:1499069 percent identity: 35.38; identified by sequence similarity; putative"
			/codon_start=1
			/transl_table=11
			/product-"conserved hypothetical protein"
			/protein_id-"AAB89785.1"



```
/db_xref="GI:2649106"
/translation="WTRDRLEIYDIILSIIRKHGNSILPTLLRFSNLSQNFNRYM
DLVKVGLVREIQDSGRYISLTDKGFRLEKYQIREFIEFGL"
gene
complement(5887..6570)
/genes="AF1460"
CDS
complement(5887..6570)
/genes="AF1460"
/notes="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1
/transl_table=11
/product="A. fulgidus predicted coding region AF1460"
/protein_id="AAB89791.1"
/db_xref="GI:2649112"
/translation="MPPMIPPKLPPEFIAEAALYSYIIIFLCIYHRLREVYKLS
DYRGHFSNTFLGLFLVFLVLLSASGWFEEISLEGRLGIMAFSMAFLAYS
SAALIVYISLAWRLERPPGEWINGVALVIALTSLSRMPLVFLISQALAVFLVA
AIFVYSHRHSRSKRAYPLIILFVFWLLNISITFRPLPLEFRFAIYTLISVAVILI
IAYRVLRKL"
gene
6658..6891
/genes="AF1461"
CDS
6658..6891
/genes="AF1461"
/notes="similar to GP:1747410 percent identity: 50.98;
identified by sequence similarity; putative"
/codon_start=1
/codon_table=11
/product="iron-sulfur cluster binding protein, putative"
/protein_id="AAB89787.1"
/db_xref="GI:2649108"
/translation="MRLGGDSVVVEIQSGICGCTCVCVEPTDVRLEGGRAVIKYP
EDCQICHLCLRLXCPVDATITSPKSLPVIVANG"
gene
7028..7579
/genes="AF1462"
CDS
7028..7579
/genes="AF1462"
/notes="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1
/transl_table=11
/product="A. fulgidus predicted coding region AF1462"
/protein_id="AAB89790.1"
/db_xref="GI:2649111"
/translation="MTVJSFTLMVNOFYLTRFRIIEGGRMSYIVKAHRLFGYISVG
FLFHPLELVPEQESQPEDAFWKIITDTSAILLIGIAGWYMLTAVTSVLRKR
LFKYNKRTFHGILAVLITVGVYLVLRHSSLANKAFAVALLAGGYVTMTKTYV
FOLRREKWKNGRLAEDSSSP"
gene
7531..9258
Query Match 2.5%; Score 20; DB 2; Length 10782;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 368 cctctcagccacctccaga 387
|||||
Db 6211 CCCTCCAGCCACCTCCAGA 6230

RESULT 10
AC014435
LOCUS
DEFINITION
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION
AC014435
VERSION
AC014435.1 GI:6436900
KEYWORDS
HTG: HTGS_PHASE2.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 22166)
REFERENCE
AUTHORS
Adams, M. and Venter, J. C.
```

```
Direct Submission
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT
This sequence was identified as CDM:10213130 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers
1..22166
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 6622 a 4486 c 4544 g 6514 t
ORIGIN
QY 114 ctggctctgctgctgctct 133
|||||
Db 14740 CCGGCTCTGCTGCTGCTCT 14759

RESULT 11
AC011166
LOCUS
DEFINITION
Homo sapiens clone 10_C_2, LOW-PASS SEQUENCE SAMPLING.
ACCESSION
AC011166
VERSION
AC011166.1 GI:6006136
KEYWORDS
HTG: HTGS_PHASE0.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 34871)
AUTHORS
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE
Homo sapiens, clone 10_C_2
JOURNAL
Unpublished
AUTHORS
2 (bases 1 to 34871)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Becker, R., Boguslavsky, L., Bouckhgalter, B.,
Brown, A., Castle, A., Collangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Horton, L.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Klein, J.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thoman, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT
All repeats were identified using RepeatMasker: Smit, A. F. A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
* NOTE: This record contains 39 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
```

\* 1 890: contig of 890 bp in length  
 \* gap of unknown length  
 \* 891 1763: contig of 873 bp in length  
 \* gap of unknown length  
 \* 1764 2669: contig of 906 bp in length  
 \* gap of unknown length  
 \* 2670 3566: contig of 897 bp in length  
 \* gap of unknown length  
 \* 3567 4428: contig of 862 bp in length  
 \* gap of unknown length  
 \* 4429 5265: contig of 837 bp in length  
 \* gap of unknown length  
 \* 5266 6127: contig of 862 bp in length  
 \* gap of unknown length  
 \* 6128 7022: contig of 895 bp in length  
 \* gap of unknown length  
 \* 7023 7913: contig of 891 bp in length  
 \* gap of unknown length  
 \* 7914 8807: contig of 894 bp in length  
 \* gap of unknown length  
 \* 8808 9679: contig of 872 bp in length  
 \* gap of unknown length  
 \* 9680 10532: contig of 853 bp in length  
 \* gap of unknown length  
 \* 10533 11406: contig of 874 bp in length  
 \* gap of unknown length  
 \* 11407 12286: contig of 880 bp in length  
 \* gap of unknown length  
 \* 12287 13172: contig of 886 bp in length  
 \* gap of unknown length  
 \* 13173 14099: contig of 927 bp in length  
 \* gap of unknown length  
 \* 14100 15011: contig of 912 bp in length  
 \* gap of unknown length  
 \* 15012 15897: contig of 886 bp in length  
 \* gap of unknown length  
 \* 15898 16769: contig of 872 bp in length  
 \* gap of unknown length  
 \* 16770 17651: contig of 882 bp in length  
 \* gap of unknown length  
 \* 17652 18654: contig of 1003 bp in length  
 \* gap of unknown length  
 \* 18655 19572: contig of 918 bp in length  
 \* gap of unknown length  
 \* 19573 20482: contig of 910 bp in length  
 \* gap of unknown length  
 \* 20483 21441: contig of 959 bp in length  
 \* gap of unknown length  
 \* 21442 22350: contig of 909 bp in length  
 \* gap of unknown length  
 \* 22351 23227: contig of 877 bp in length  
 \* gap of unknown length  
 \* 23228 24104: contig of 877 bp in length  
 \* gap of unknown length  
 \* 24105 25111: contig of 1007 bp in length  
 \* gap of unknown length  
 \* 25112 26005: contig of 894 bp in length  
 \* gap of unknown length  
 \* 26006 26888: contig of 883 bp in length  
 \* gap of unknown length  
 \* 26889 27763: contig of 875 bp in length  
 \* gap of unknown length  
 \* 27764 28721: contig of 958 bp in length  
 \* gap of unknown length  
 \* 28722 29587: contig of 866 bp in length  
 \* gap of unknown length  
 \* 29588 30440: contig of 853 bp in length  
 \* gap of unknown length  
 \* 30441 31313: contig of 873 bp in length  
 \* gap of unknown length  
 \* 31314 32224: contig of 911 bp in length  
 \* gap of unknown length  
 \* 32225 33101: contig of 877 bp in length  
 \* gap of unknown length

\* 33102 33996: gap of unknown length  
 \* contig of 895 bp in length  
 \* 33997 34871: contig of 875 bp in length.  
 \* Location/Qualifiers  
 \* 1. 34871  
 \* /organism="Homo sapiens"  
 \* /db\_xref="taxon:9606"  
 \* /clone="10\_C.2"  
 \* /clone.lib="RPCI-11 Human Male BAC"  
 \* BASE COUNT 8119 a 9192 c 7777 g 9233 t 550 others  
 \* ORIGIN  
 \* Query Match 2.5%: Score 20; DB 43; Length 34871;  
 \* Best Local Similarity 100.0%; Pred. No. 2.7;  
 \* Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 \* QY 287 ctctgccccctggacttcc 306  
 \* |||||  
 \* Db 3752 CTCTGCCCATGGACTTCC 3771  
 \* RESULT 12  
 \* AC010844  
 \* LOCUS  
 \* DEFINITION  
 \* Drosophila melanogaster chromosome X clone BACR03F03 (D882) RPCI-98  
 \* 03.F.3 map 12C-12D strain Y; cn bw sp, \*\*\* SEQUENCING IN PROGRESS  
 \* \*\* 100 unordered pieces.  
 \* AC010844  
 \* AC010844.4 GI:6016642  
 \* HTG: HTGS\_PHASE1.  
 \* fruit fly.  
 \* ORGANISM  
 \* Drosophila melanogaster  
 \* Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 \* Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 \* Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 \* 1 (bases 1 to 92509)  
 \* Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,  
 \* Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
 \* Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
 \* Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
 \* Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
 \* Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,  
 \* Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,  
 \* Svirskaas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
 \* Rubin, G.M.  
 \* TITLE  
 \* Sequencing of Drosophila melanogaster  
 \* JOURNAL  
 \* Unpublished  
 \* REFERENCE  
 \* 2 (bases 1 to 92509)  
 \* Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,  
 \* Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
 \* Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
 \* Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
 \* Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
 \* Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,  
 \* Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,  
 \* Svirskaas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
 \* Rubin, G.M.  
 \* TITLE  
 \* Direct Submission  
 \* JOURNAL  
 \* Submitted (24-SEP-1999) Drosophila Genome Center, Lawrence Berkeley  
 \* Laboratory, MS 64-121, Berkeley, CA 94720, USA  
 \* On Oct 8, 1999 this sequence version replaced gi:6006066.  
 \* For further information about this sequence, including its location  
 \* and relationship to other sequences, please visit our sequence  
 \* archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
 \* to [bdgs@fruitfly.berkeley.edu](mailto:bdgs@fruitfly.berkeley.edu). All contigs in this submission meet  
 \* the following cutoffs: length >= 200 bases.  
 \* \* NOTE: This is a 'working draft' sequence. It currently  
 \* \* consists of 100 contigs. The true order of the pieces  
 \* \* is not known and their order in this sequence record is  
 \* \* arbitrary. Gaps between the contigs are represented as  
 \* \* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 664: contig of 664 bp in length  
\* 665 744: gap of unknown length  
\* 745 1298: contig of 554 bp in length  
\* 1299 1378: gap of unknown length  
\* 1379 1987: contig of 609 bp in length  
\* 1988 2067: gap of unknown length  
\* 2068 2719: contig of 652 bp in length  
\* 2720 2799: gap of unknown length  
\* 2800 3032: contig of 1133 bp in length  
\* 3033 4012: gap of unknown length  
\* 4013 4712: contig of 700 bp in length  
\* 4713 4792: gap of unknown length  
\* 4793 5546: contig of 754 bp in length  
\* 5547 5626: gap of unknown length  
\* 5627 6428: contig of 802 bp in length  
\* 6429 6508: gap of unknown length  
\* 6509 7234: contig of 726 bp in length  
\* 7235 7314: gap of unknown length  
\* 7315 8210: contig of 896 bp in length  
\* 8211 8290: gap of unknown length  
\* 8291 8957: contig of 667 bp in length  
\* 8958 9307: gap of unknown length  
\* 9309 9340: contig of 903 bp in length  
\* 9341 10020: gap of unknown length  
\* 10021 10867: contig of 847 bp in length  
\* 10868 10947: gap of unknown length  
\* 10948 11592: contig of 645 bp in length  
\* 11593 11672: gap of unknown length  
\* 11673 12300: contig of 628 bp in length  
\* 12301 12380: gap of unknown length  
\* 12381 12785: contig of 405 bp in length  
\* 12786 12865: gap of unknown length  
\* 12866 13713: contig of 848 bp in length  
\* 13714 13793: gap of unknown length  
\* 13794 14731: contig of 938 bp in length  
\* 14732 14811: gap of unknown length  
\* 14812 15574: contig of 763 bp in length  
\* 15575 15654: gap of unknown length  
\* 15655 16762: contig of 1108 bp in length  
\* 16763 16842: gap of unknown length  
\* 16843 17720: contig of 878 bp in length  
\* 17721 17800: gap of unknown length  
\* 17801 18710: contig of 910 bp in length  
\* 18711 18790: gap of unknown length  
\* 18791 19193: contig of 403 bp in length  
\* 19194 19273: gap of unknown length  
\* 19274 20395: contig of 922 bp in length  
\* 20396 20375: gap of unknown length  
\* 20376 21358: contig of 1083 bp in length  
\* 21359 21438: gap of unknown length  
\* 21439 22656: contig of 1218 bp in length  
\* 22657 22736: gap of unknown length  
\* 22737 23743: contig of 1007 bp in length  
\* 23744 23823: gap of unknown length  
\* 23824 25367: contig of 1544 bp in length  
\* 25368 25447: gap of unknown length  
\* 25448 26594: contig of 1147 bp in length  
\* 26595 26874: gap of unknown length  
\* 26875 27986: contig of 1312 bp in length  
\* 27987 28066: gap of unknown length  
\* 28067 28982: contig of 916 bp in length  
\* 28983 29062: gap of unknown length  
\* 29063 29924: contig of 862 bp in length  
\* 29925 30004: gap of unknown length  
\* 30005 30968: contig of 964 bp in length  
\* 30969 31048: gap of unknown length  
\* 31049 32547: contig of 1499 bp in length  
\* 32548 32627: gap of unknown length  
\* 32628 33808: contig of 1181 bp in length  
\* 33809 33888: gap of unknown length

\* 33889 34986: contig of 1098 bp in length  
\* 34987 35066: gap of unknown length  
\* 35067 35803: contig of 737 bp in length  
\* 35804 35883: gap of unknown length  
\* 35884 36294: contig of 411 bp in length  
\* 36295 36374: gap of unknown length  
\* 36375 37470: contig of 1096 bp in length  
\* 37471 37550: gap of unknown length  
\* 37551 39563: contig of 2013 bp in length  
\* 39564 39643: gap of unknown length  
\* 39644 41050: contig of 1407 bp in length  
\* 41051 41130: gap of unknown length  
\* 41131 42389: contig of 1259 bp in length  
\* 42390 42469: gap of unknown length  
\* 42470 44106: contig of 1637 bp in length  
\* 44107 44186: gap of unknown length  
\* 44187 45257: contig of 1071 bp in length  
\* 45258 45337: gap of unknown length  
\* 45338 48008: contig of 1471 bp in length  
\* 48009 48888: gap of unknown length  
\* 48889 49104: contig of 2216 bp in length  
\* 49105 49184: gap of unknown length  
\* 49185 50451: contig of 1267 bp in length  
\* 50452 50531: gap of unknown length  
\* 50532 52592: contig of 2061 bp in length  
\* 52593 52672: gap of unknown length  
\* 52673 55002: contig of 2330 bp in length  
\* 55003 55082: gap of unknown length  
\* 55083 57023: contig of 1941 bp in length  
\* 57024 57103: gap of unknown length  
\* 57104 59408: contig of 2305 bp in length  
\* 59409 59488: gap of unknown length  
\* 59489 60015: contig of 527 bp in length  
\* 60016 60095: gap of unknown length  
\* 60096 60452: contig of 357 bp in length  
\* 60453 60532: gap of unknown length  
\* 60533 61086: contig of 554 bp in length  
\* 61087 61166: gap of unknown length  
\* 61167 61961: contig of 795 bp in length  
\* 61962 62041: gap of unknown length  
\* 62042 62577: contig of 536 bp in length  
\* 62578 62657: gap of unknown length  
\* 62658 63205: contig of 548 bp in length  
\* 63206 63285: gap of unknown length  
\* 63286 63767: contig of 482 bp in length  
\* 63768 63847: gap of unknown length  
\* 63848 64432: contig of 585 bp in length  
\* 64433 64512: gap of unknown length  
\* 64513 65290: contig of 778 bp in length  
\* 65291 65370: gap of unknown length  
\* 65371 66186: contig of 816 bp in length  
\* 66187 66266: gap of unknown length  
\* 66267 66724: contig of 458 bp in length  
\* 66725 66804: gap of unknown length  
\* 66805 67422: contig of 618 bp in length  
\* 67423 67502: gap of unknown length  
\* 67503 68287: contig of 785 bp in length  
\* 68288 68367: gap of unknown length  
\* 68368 68994: contig of 627 bp in length  
\* 68995 69074: gap of unknown length  
\* 69075 69731: contig of 657 bp in length  
\* 69732 69811: gap of unknown length  
\* 69812 70509: contig of 698 bp in length  
\* 70510 70589: gap of unknown length  
\* 70590 71259: contig of 670 bp in length  
\* 71260 71339: gap of unknown length  
\* 71340 71975: contig of 636 bp in length  
\* 71976 72055: gap of unknown length  
\* 72056 72256: contig of 201 bp in length  
\* 72257 72336: gap of unknown length  
\* 72337 73239: contig of 903 bp in length  
\* 73240 73319: gap of unknown length  
\* 73320 73974: contig of 655 bp in length

```

* 73975 74054: gap of unknown length
* 74055 74743: contig of 689 bp in length
* 74744 74823: gap of unknown length
* 74824 75343: contig of 520 bp in length
* 75344 75423: gap of unknown length
* 75424 76021: contig of 598 bp in length

Query Match      2.5%  Score 20; DB 41; Length 92509;
Best Local Similarity 100.0%; Pred. No. 2.7; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

Qy 114 ctgctctgtctgtctctct 133
|||||
Db 82726 CTGCTCTGCTGCTGCTCT 82745

RESULT 13
AC016738
LOCUS
DEFINITION Homo sapiens clone RP11-426K3, WORKING DRAFT SEQUENCE, 1 unordered
pieces.
ACCESSION AC016738
VERSION AC016738.1 GI:6524317
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 114996)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 114996)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT Center project name: H_NH0426K03.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 114996: contig of 114996 bp in length.
FEATURES
source
1..114996
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-426K3"
BASE COUNT 31274 a 25793 c 25551 g 31748 t 630 others
ORIGIN

Query Match      2.5%  Score 20; DB 44; Length 114996;
Best Local Similarity 100.0%; Pred. No. 2.7; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

Qy 369 ctctccagccacctccagat 388
|||||
Db 15109 CTCTCCAGCCACCTCCAGAT 15128

RESULT 14
AC007099
LOCUS
DEFINITION Homo sapiens BAC clone NH0445A14 from 2, complete sequence.
ACCESSION AC007099
VERSION AC007099.3 GI:4753234
KEYWORDS HTG.

```

---

```

SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 125032)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 125032)
AUTHORS Kozlowicz,A., Wohlmann,P., Hawkins,M. and Sapetti,L.
TITLE The sequence of Homo sapiens BAC clone NH0445A14
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 125032)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (16-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 125032)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 125032)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On May 5, 1999 this sequence version replaced g1:4454615.
SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108 USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is NH0342K06, 200 bp overlap.
Actual start of this clone is at base position 163358 of NH0342K06;
actual end is at base position 125032 of NH0445A14.
FEATURES
Location/Qualifiers

```

```

source
1. .125032
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="NH0445A14"
/clone_lib="RPCI-11"
645. .5978
/rpt_family="L1"
5979. .6282
/rpt_family="Alu"
6283. .8556
/rpt_family="L1"
8567. .8610
/rpt_family="AT_rich"
8624. .8818
/rpt_family="MaLR"
8830. .9102
/rpt_family="L1"
9103. .9259
/rpt_family="MaLR"
9475. .10091
/rpt_family="L1"
10339. .10755
/rpt_family="L1"
10756. .11066
/rpt_family="Alu"
11067. .11100
/rpt_family="L1"
11178. .11303
/rpt_family="MIR"
11266. .11314
/rpt_family="L2"
11659. .12178
/rpt_family="Other"
13272. .13812
/rpt_family="MaLR"
14085. .14276
/rpt_family="MIR"
14349. .14830
/rpt_family="MaLR"
15286. .15459
/rpt_family="MIR"
17284. .17761
/rpt_family="L2"
17733. .18118
/rpt_family="L2"
18119. .18160
/rpt_family="AT_rich"
18161. .18841
/rpt_family="L1"
18842. .18857
/rpt_family="AT_rich"
18895. .19002
/rpt_family="(TTAA)n"
20118. .20163
/rpt_family="AT_rich"
21298. .21346
/rpt_family="(GA)n"
21977. .22136
/rpt_family="MIR"
22400. .22442
/rpt_family="MIR"
23485. .23532
/rpt_family="MIR"
23675. .23804
/rpt_family="MIR"
24095. .24239
/rpt_family="MIR"
24414. .24475
/rpt_family="L2"
24669. .24735
/rpt_family="MIR"

```

```

repeat_region 24923. .25089
/rpt_family="MERL_type"
repeat_region 25152. .25306
/rpt_family="MERL_type"
misc_feature 25988. .26054
/notes="match to EST AA234473 (NID:g1858965) zr74a03.rl"
misc_feature 25988. .26054
/notes="similar to EST AA186620 (NID:gl774719) zo71h03.rl"
misc_feature 25993. .26054
/notes="similar to EST H64645 (NID:g1023385) yu63c03.rl"
misc_feature 25998. .26054
/notes="match to EST AA526962 (NID:g2269031) n106b07.sl"
misc_feature 26346. .26637
/notes="match to EST W25216 (NID:gl303119) zb87a10.rl"
misc_feature 26392. .26631
/notes="match to EST AA234473 (NID:g1858965) zr74a03.rl"
misc_feature 26392. .26607
/notes="similar to EST H64645 (NID:g1023385) yu63c03.rl"
misc_feature 26392. .26580
/notes="similar to EST AA186620 (NID:gl774719) zo71h03.rl"
misc_feature 26392. .26631
/notes="match to EST AA526962 (NID:g2269031) n106b07.sl"
misc_feature 26395. .26631
/notes="match to EST AI041947 (NID:g3281141) oy10e12.xl"
misc_feature 26466. .26622
/notes="match to EST N45383 (NID:gl186549) yw97c05.rl"
repeat_region 27386. .27520
/rpt_family="MIR"
repeat_region 27842. .27958
/rpt_family="MIR"
repeat_region 27974. .28208
/rpt_family="MERL_type"
repeat_region 28408. .28487
/rpt_family="MERL_type"
misc_feature 28612. .28737
/notes="match to EST AA234473 (NID:g1858965) zr74a03.rl"
misc_feature 28612. .28828
/notes="match to EST AI041947 (NID:g3281141) oy10e12.xl"
misc_feature 28612. .28828
/notes="match to EST AA526962 (NID:g2269031) n106b07.sl"
repeat_region 28796. .28816
/rpt_family="AT_rich"
repeat_region 28906. .29009
/rpt_family="MIR"
repeat_region 29073. .29331
/rpt_family="MIR"
repeat_region 29338. .29383
/rpt_family="L2"
repeat_region 30372. .30654

```

Query Match 2.5%; Score 20; DB 40; Length 125032;  
Best Local Similarity 100.0%; Pred.No. 2.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 763 ccataatggaataaagtcttct 782  
|||||  
Db 10395 CCATATGGAATAAAGTCT 10414

# RESULT 15

AC009647  
LOCUS AC009647 160848 bp DNA HTG 05-NOV-1999  
DEFINITION Homo sapiens chromosome 11 clone 164\_L\_18 map 11, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 15 unordered pieces.  
ACCESSION AC009647  
VERSION AC009647.2 GI:6249733  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE Human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 160848)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Homo sapiens chromosome 11, clone 164\_L\_18  
Unpublished  
2 (bases 1 to 160848)  
Birken,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,  
Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,  
Cooke,P., DeArelilano,K., Depayre,E., Devon,K., Dewar,K.,  
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,  
Funke,R., Gage,D., Gargan,J., Gardyna,S., Gilbert,D., Grant,G.,  
Hagos,B., Hearford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,  
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,  
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,  
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,  
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,  
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,  
Tessaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,  
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

Direct Submission  
Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Nov 5, 1999 this sequence version replaced gi:5801761.  
All repeats were identified using RepeatMasker: Smit, A.F.A. &  
Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 15 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

TITLE  
JOURNAL  
COMMENT

1 3081: contig of 3081 bp in length  
\*  
\* 3082 6103: contig of 3022 bp in length  
\*  
\* 6104 10175: contig of 4072 bp in length  
\*  
\* 10176 17283: contig of 7108 bp in length  
\*  
\* 17284 23742: contig of 6459 bp in length  
\*  
\* 23743 30929: contig of 7187 bp in length  
\*  
\* 30930 38194: contig of 7265 bp in length  
\*  
\* 38195 48623: contig of 10429 bp in length  
\*  
\* 48624 59508: contig of 10885 bp in length  
\*  
\* 59509 68964: contig of 9456 bp in length  
\*  
\* 68965 79909: contig of 10945 bp in length  
\*  
\* 79910 93341: contig of 13432 bp in length  
\*  
\* 93342 114403: contig of 21062 bp in length  
\*  
\* 114404 138709: contig of 24306 bp in length  
\*  
\* 138710 160848: contig of 22139 bp in length.

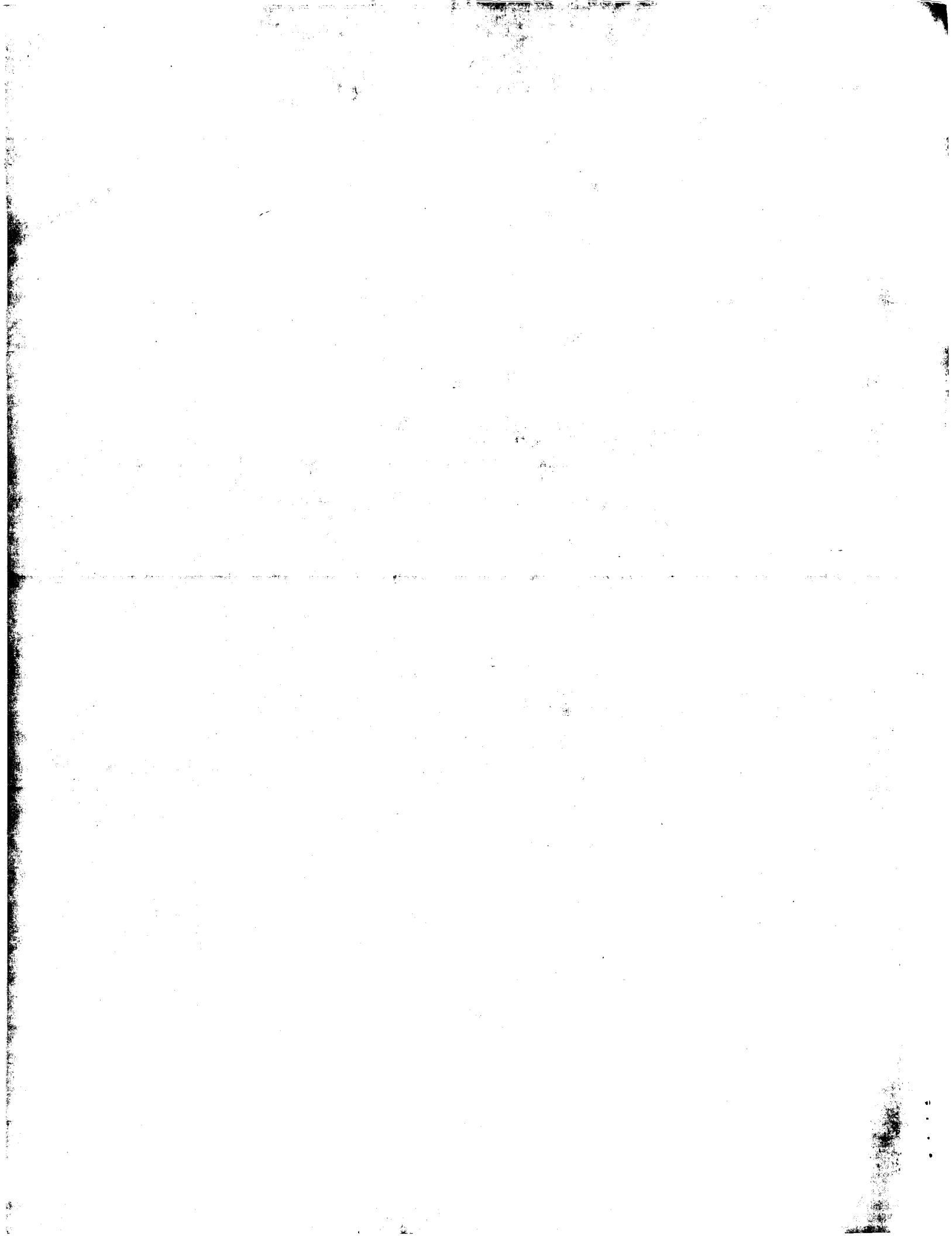
FEATURES  
source

1. 160848  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11"  
/clone="164\_L\_18"  
/clone\_lib="RPC1-11 Human Male BAC"  
36905 a 43772 c 43253 g 36893 t 25 others  
BASE COUNT  
ORIGIN

Query Match 2.5%; Score 20; DB 43; Length 160848;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 118 tcctgctgctgctcctctctct 137  
|||||  
Db 44868 TCCTGCTGCTGCTCCTCTCT 44887

Search completed: May 1, 2000, 16:47:50  
Job time: 20970 sec

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 18:53:48 ; Search time 250.34 Seconds  
(without alignments)  
784.536 Million cell updates/sec

Title: US-09-215-435-118  
Perfect score: 785  
Sequence: 1 cggaaatccggagtcggtg.....tatggaaataaagttcttc 785

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 311585 seqs, 125096042 residues

Word size : 0

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	393	50.1	928	1 V07910	Human cysteine-rich
2	325	41.4	464	1 X51459	Human secreted pro
3	18	2.3	1024	1 Q23904	Rat GTP cyclohydro
C 4	18	2.3	1033	1 X01911	Human HECOP cDNA f
C 5	18	2.3	1210	1 V84425	Human secreted pro
6	18	2.3	2228	1 Q98986	2B10 AMH-receptor
7	18	2.3	2408	1 Q98987	H1 AMH-receptor cD
C 8	18	2.3	3663	1 T11342	nod gene fragment
9	18	2.3	3736	1 X29464	Human negative reg
10	18	2.3	4558	1 Q43661	Acetobacter cdg2 o
C 11	18	2.3	5294	1 T89253	CDNA for human CAS
12	17	2.2	39	1 T40262	Human ob protein f
13	17	2.2	39	1 T51178	Human leptin PCR f
14	17	2.2	69	1 Q36042	Chicken nov gene e
15	17	2.2	231	1 T10683	Mucin-derived prot
16	17	2.2	258	1 T10684	Mucin-derived prot
17	17	2.2	291	1 T10685	Mucin-derived prot
18	17	2.2	318	1 T10686	Mucin-derived prot
19	17	2.2	459	1 Q12772	Human polymorphic
20	17	2.2	558	1 Q24678	H23-ETA-T antigen
21	17	2.2	558	1 Q24680	H23-ETA-S antigen
22	17	2.2	723	1 T10681	Mucin-derived prot
23	17	2.2	750	1 T10682	Mucin-derived prot
24	17	2.2	768	1 T10679	Mucin-derived prot
25	17	2.2	795	1 T10680	Mucin-derived prot
26	17	2.2	822	1 T10677	Mucin-derived prot
27	17	2.2	849	1 T10678	Mucin-derived prot
28	17	2.2	1109	1 Q03852	Sequence encoding
29	17	2.2	1109	1 Q20612	Murine macrophage
30	17	2.2	1109	1 Q20728	Murine macrophage
C 31	17	2.2	1149	1 T00737	Multiple tumour su
C 32	17	2.2	1149	1 Q98166	Human MTS1 DNA inc
C 33	17	2.2	1149	1 V11239	Human MTS1 genomic

c 34	17	2.2	1149	1 V53820	Coding sequence 1
c 35	17	2.2	1149	1 V70584	Human multiple tum
36	17	2.2	1452	1 Q29276	Encodes transmembr
37	17	2.2	1527	1 V48329	Minimuc1 gene. New
38	17	2.2	1567	1 Q12773	Human polymorphic
39	17	2.2	1709	1 Q29277	Encodes secreted f
40	17	2.2	1763	1 Q12771	Human polymorphic
41	17	2.2	1785	1 V15493	Aeromonas caviae p
42	17	2.2	1787	1 V52941	Human UNC-5 homolo
c 43	17	2.2	1884	1 V17814	Homo sapiens don-1
c 44	17	2.2	1900	1 X29139	Hypoxia-regulated
45	17	2.2	1975	1 Q36031	Chicken nov coding

## ALIGNMENTS

RESULT 1	
V07910	
ID	V07910 standard; cDNA; 928 BP.
AC	V07910;
DE	18-JAN-1999 (first entry)
DT	Human cysteine-rich secreted protein-like-N cDNA.
DE	Human cysteine-rich secreted protein; tumour; cancer;
KW	CRSP-like-N; cysteine-rich secreted protein; tumour; cancer;
KW	signal transduction; cell differentiation; cell proliferation;
KW	human; ss.
OS	Homo sapiens.
FH	Key
FT	Location/Qualifiers
CDS	75..803
FT	/*tag= a
PN	W09846755-A1.
PD	22-OCT-1998.
PF	16-APR-1998; U07894.
PR	20-JAN-1998; US-009802.
PR	16-APR-1997; US-843704.
PR	17-APR-1997; US-842898.
PR	15-JAN-1998; US-071589.
PA	(MILL-) MILLENNIUM BIOTHEAPEUTICS INC.
PI	McCarthy SA;
DR	WPI; 98-568730/48.
DR	P-PSDB; W73020.
PT	New isolated cysteine-rich secreted proteins - used to develop
PT	products for treating, e.g. hyperproliferative disorders, cancers,
PT	wounds, infectious lesions, degenerative lesions or demyelating
PT	diseases
PS	Example 2; Page 107-108; 142pp; English.
CC	This full-length cDNA clone codes for a novel human protein (see
CC	W73020), designated CRSP-like-N, that shows homology to a novel,
CC	claimed human cysteine-rich secreted protein 1 (CRSP-1, see
CC	W73016). CRSP-like-N was identified using the nucleotide
CC	sequence encoding the N-terminal unique region of CRSP-1 to
CC	search a protein sequence database. Clone AA397836 was identified
CC	and was subsequently isolated from the IMAGE collection and
CC	sequenced fully to define the entire hCRSP-like-n sequence.
CC	CRSP-1, -2, -3 and -4 nucleic acid sequences (see V07906-09) and
CC	polypeptides (see W73016-19) are claimed. These can be used in
CC	diagnostic, screening and therapeutic methods of the invention e.g.
CC	for treating hyperproliferative disorders, cancers, wounds, and
CC	infectious or degenerative lesions and demyelating diseases, and in
CC	drug screening.
SQ	Sequence 928 BP; 229 A; 291 C; 257 G; 151 T;

Query Match	50.1%;	Score 393;	DB 1;	Length 928;
Best Local Similarity	100.0%;	Pred. No. 2.5e-175;		
Matches 393;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	393	aaggtaccaggatgagagaggagggccctgggtaccatccagagccagcgacagc	452	
Db	489	AAGGTACCAGGATGAGGAGGAGGAGGCGCTCGTACCATCCAGAGGCCAGGCAGC	548	
QY	453	ttccacacagaactccatccccgggtggtccttgatcatttaagctgcacggcgagg	512	

[illegible]

```

RESULT 3
Q23904
ID Q23904 standard; DNA; 1024 BP.
AC Q23904;
DE 13-OCT-1992 (first entry)
DE Rat GTP cyclohydrolase-I coding sequence.
KW guanosine-5'-triphosphate; monoamine neurotransmitter; deficiency;
KW pterin biosynthesis; ss.
OS Rattus.
EH Key Location/Qualifiers
FT cds 128..853
FT /*tag= a
FT /product= GTP_cyclohydrolase-I
FT
FN J04082888-A.
PD 16-MAR-1992.
PF 21-JUL-1990; 193359.
PR 21-JUL-1990: JP-193359.
PA (SUNR ) SUNTORY LTD.
DR WPI; 92-138663/17.
DR P-PSDB; R22903.
DR Tetrahydrobiopterin prepn. - used to treat deficiencies in
PT mono-amine neuro-transmitter by reacting GTP with GTP
PT cyclo:hydrolase, 6-pyruvoyl synthase and reductase
PS Disclosure; Fig 10; 23pp; Japanese.
CC This sequence codes for the rat GTP cyclohydrolase I. The inventors
CC have partially characterised the GTP cyclohydrolase-I from E.coli
CC which converts GTP into D-erythro-7,8-dihydroneopterin triphosphate.
CC See also R22864-6 and Q24164-8.
SQ Sequence 1024 BP; 244 A; 279 C; 291 G; 210 T;

Query Match 2.3%; Score 18; DB 1; Length 1024;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

20-222, 1999 (first entry)  
Human HECOP CDNA from library COLNFET02.  
HECOP; human coatome vesicle protein; treatment; diabetes; AIDS; viral;  
abnormal vesicle trafficking; gastrointestinal disorder; allergy; fungal;  
autoimmune haemolytic anaemia; multiple sclerosis; myasthenia gravis;  
arthritis; scleroderma; Chediak-Higashi syndrome; Sjogren's syndrome;  
abnormal vesicle trafficking; gastrointestinal disorder; allergy; fungal;  
autoimmune haemolytic anaemia; multiple sclerosis; myasthenia gravis;  
arthritis; scleroderma; Chediak-Higashi syndrome; Sjogren's syndrome;

KW systemic lupus erythematosus; toxic shock syndrome; infection; bacterial;  
OS helminth; protozoa; ss.  
FH Homo sapiens.  
FT Key Location/Qualifiers  
CDS 14..940  
FT /\*tag= a  
FT /product= "HECOP"  
FT /note= "human coatamer vesicle protein"  
FN US5869256-A.  
PD 09-FEB-1999.  
PF 24-FEB-1997; 807050.  
PR 24-FEB-1997; US-807050.  
PA (INCY-) INCYTE PHARM INC.  
PI Goli SK, Hillman JL;  
DR WPI; 99-152759/13.  
DR P-PSDB; W92405.  
PT Human coatamer vesicle protein HECOP - useful for treating disorders  
PT associated with abnormal vesicle trafficking  
PS Claim 4; Fig 1A-B; 2ppp; English.  
CC This sequence encodes a novel human coatamer vesicle protein (HECOP)  
CC isolated from cDNA library COLNFT02. The encoded polypeptide or vectors  
CC expressing it, can be used to treat disorders associated with abnormal  
CC vesicle trafficking, e.g. cystic fibrosis, glucose-galactose  
CC malabsorption syndrome, hypercholesterolaemia, diabetes mellitus,  
CC diabetes insipidus, hyper- and hypoglycaemia, Grave's disease, goitre,  
CC Cushing's disease, Addison's disease, gastrointestinal disorders  
CC including ulcerative colitis, gastric and duodenal ulcers, AIDS,  
CC allergies including hay fever, asthma and urticaria (hives), autoimmune  
CC haemolytic anaemia, proliferative glomerulonephritis, inflammatory bowel  
CC disease, multiple sclerosis, myasthenia gravis, rheumatoid and  
CC osteoarthritis, scleroderma, Chediak-Higashi and Sjogren's syndromes,  
CC systemic lupus erythematosus, toxic shock syndrome, traumatic tissue age  
CC and viral, bacterial, fungal, helminth and protozoal infections.  
SQ Sequence 1033 BP; 221 A; 328 C; 319 G; 160 T;

Query Match 2.3%; Score 18; DB 1; Length 1033;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 642 cactcaggctgtccccc 659  
|||||  
Db 456 CACTCCAGGCTGCCCC 439

RESULT 5  
V84425/c  
ID V84425 standard; DNA; 1210 BP.  
AC V84425.  
DT 01-MAR-1999 (first entry)  
DE Human secreted protein gene 15 clone HPTBB03.  
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
OS Homo sapiens.  
FN W03854963-A2.  
PD 10-DEC-1998.  
PF 04-JUN-1998; U11422.  
PR 18-DEC-1997; US-070923.  
PR 06-JUN-1997; US-048877.  
PR 06-JUN-1997; US-048881.  
PR 06-JUN-1997; US-048884.  
PR 06-JUN-1997; US-048893.  
PR 06-JUN-1997; US-048896.  
PR 06-JUN-1997; US-048899.  
PR 06-JUN-1997; US-048915.  
PR 06-JUN-1997; US-048949.  
PR 06-JUN-1997; US-048964.

PR 06-JUN-1997; US-048972.  
PR 06-JUN-1997; US-049020.  
PR 06-JUN-1997; US-049375.  
PR 05-SEP-1997; US-057628.  
PR 05-SEP-1997; US-057635.  
PR 05-SEP-1997; US-057644.  
PR 05-SEP-1997; US-057647.  
PR 05-SEP-1997; US-057650.  
PR 05-SEP-1997; US-057661.  
PR 05-SEP-1997; US-057667.  
PR 05-SEP-1997; US-057671.  
PR 05-SEP-1997; US-057764.  
PR 05-SEP-1997; US-057770.  
PR 05-SEP-1997; US-057775.  
PR 05-SEP-1997; US-057778.  
PR 06-JUN-1997; US-048875.  
PR 06-JUN-1997; US-048878.  
PR 06-JUN-1997; US-048882.  
PR 06-JUN-1997; US-048885.  
PR 06-JUN-1997; US-048894.  
PR 06-JUN-1997; US-048897.  
PR 06-JUN-1997; US-048900.  
PR 06-JUN-1997; US-048916.  
PR 06-JUN-1997; US-048962.  
PR 06-JUN-1997; US-048970.  
PR 06-JUN-1997; US-048974.  
PR 06-JUN-1997; US-049373.  
PR 05-SEP-1997; US-057584.  
PR 05-SEP-1997; US-057629.  
PR 05-SEP-1997; US-057642.  
PR 05-SEP-1997; US-057645.  
PR 05-SEP-1997; US-057648.  
PR 05-SEP-1997; US-057651.  
PR 05-SEP-1997; US-057662.  
PR 05-SEP-1997; US-057668.  
PR 05-SEP-1997; US-057762.  
PR 05-SEP-1997; US-057765.  
PR 05-SEP-1997; US-057771.  
PR 05-SEP-1997; US-057776.  
PR 06-JUN-1997; US-048876.  
PR 06-JUN-1997; US-048880.  
PR 06-JUN-1997; US-048883.  
PR 06-JUN-1997; US-048892.  
PR 06-JUN-1997; US-048895.  
PR 06-JUN-1997; US-048898.  
PR 06-JUN-1997; US-048901.  
PR 06-JUN-1997; US-048917.  
PR 06-JUN-1997; US-048963.  
PR 06-JUN-1997; US-048971.  
PR 06-JUN-1997; US-049019.  
PR 06-JUN-1997; US-049374.  
PR 05-SEP-1997; US-057627.  
PR 05-SEP-1997; US-057634.  
PR 05-SEP-1997; US-057643.  
PR 05-SEP-1997; US-057646.  
PR 05-SEP-1997; US-057649.  
PR 05-SEP-1997; US-057654.  
PR 05-SEP-1997; US-057666.  
PR 05-SEP-1997; US-057760.  
PR 05-SEP-1997; US-057763.  
PR 05-SEP-1997; US-057769.  
PR 05-SEP-1997; US-057774.  
PR 05-SEP-1997; US-057777.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,  
PI Fan P, Feng P, Ferrie AM, Fischer CL, Florence C,  
PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW,  
PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,  
PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;  
DR WPI; 99-059865/05.  
DR P-PSDB; W88548, W88758, W88759.  
PT New isolated human genes and the secreted polypeptides they encode -  
PT useful for diagnosis and treatment of e.g. cancers, neurological

PT disorders, immune diseases, inflammation or blood disorders  
 PS Claim 4; Page 281-282; 772pp; English.  
 CC The invention relates to nucleic acid sequences (V84411 to V84633)  
 CC encoding human secreted proteins (W88534 to W88756). The secreted protein  
 CC gene sequences are deposited with the ATCC under deposit numbers ATCC  
 CC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,  
 CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host  
 CC cells comprising recombinant vectors containing the nucleic acid  
 CC sequences are used for the recombinant production of the secreted  
 CC proteins. The polynucleotide and amino acid sequences are useful for are  
 CC useful for preventing, treating or ameliorating medical conditions e.g.  
 CC by protein or gene therapy. Pathological conditions can be also diagnosed  
 CC by determining the amount of the new polypeptides in a sample or by  
 CC determining the presence of mutations in the new polynucleotides.  
 CC Specific uses are described for each of the polynucleotides, based on  
 CC which tissues they are most highly expressed in, and include developing  
 CC products for the diagnosis or treatment of cancer, neurodegenerative  
 CC disorders, developmental abnormalities and foetal deficiencies, blood  
 CC disorders, tumours, leukemias, diseases of the immune system, autoimmune  
 CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,  
 CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,  
 CC restenosis, prostate diseases, obesity, disorders involving osteoclasts  
 CC such as osteoporosis, arthritis or malignancies, diseases of testes, lung  
 CC or thymus, digestive/endocrine disorders, infections and AIDS. The  
 CC polypeptides are also useful for identifying their binding partners.  
 CC The present sequence represents a gene encoding a human secreted protein  
 CC (see descriptor line for gene number and clone identification).  
 SQ Sequence 1210 BP; 290 A; 380 C; 349 G; 185 T;

Query Match 2.3%; Score 18; DB 1; Length 1210;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 642 cactccaggctgtccccc 659  
 |||||  
 Db 519 CACTCCAGGCTGTCCCCC 502

## RESULT 6

Q98986  
 ID Q98986 standard; cDNA: 2228 BP.  
 AC Q98986;  
 DT 29-FEB-1996 (first entry)  
 DE 2B10 AMH-receptor cDNA.  
 KW 2B10 anti-Mullerian hormone receptor; anti-Mullerian hormone; AMH;  
 KW receptor; antibody; therapy; H1; 2B10; tumour;  
 KW transforming growth factor-beta; TGF-beta; bone morphogenic protein;  
 KW activin; Mullerian inhibiting substance; MIS; ss.  
 OS Oryctolagus cuniculus.  
 FH Key Location/Qualifiers  
 FT cds 62..1588  
 FT /tag- a  
 FT /product- AMH receptor  
 FT signal\_peptide 62..112  
 FT /tag- b  
 FT mat\_peptide 113..1585  
 FT /tag- c  
 FT /product- AMH-receptor  
 FT misc\_feature 113..1585  
 FT /tag- d  
 FT /note- "claimed"  
 FT misc\_feature 113..310  
 FT /tag- e  
 FT /note- "extracellular domain"  
 FT misc\_feature 311..388  
 FT /tag- f  
 FT /note- "transmembrane region"  
 FT misc\_feature 389..1585  
 FT /tag- g  
 FT /note- "cytoplasmic domain"  
 FT W09516709-A2.  
 PD 22-JUN-1995.

PF 13-DEC-1994; U14643.  
 PR 13-DEC-1993; US-166333.  
 PR 23-DEC-1993; US-173512.  
 PA (BIOJ ) BIOGEN INC.  
 PA (INRM ) INSERM INST NAT SANTE & RECH MED.  
 PI Cate RL, Jossso N;  
 DR WPI: 95-231521/30;  
 DR P-PSDB; R80631.  
 PT New DNA encoding anti-Mullerian hormone receptor, related  
 PT polypeptide(s) and antibodies - useful in immunoassays screening for  
 PT ligands, treatment of cancer cells expressing the receptor etc.  
 PS Claim 1; Page 40-42; 77pp; English.  
 CC This sequence represents the cDNA encoding the 2B10 anti-Mullerian  
 CC hormone (AMH) receptor isoform. The difference between this isoform, and  
 CC the H1 isoform is that H1 contains an extra exon. AMH is a glycoprotein  
 CC and is part of the transforming growth factor-beta superfamily. AMH is  
 CC structurally and functionally related to TGF-beta bone morphogenic  
 CC protein, and activin. AMH receptors are present in a limited number of  
 CC tissues and can therefore be used to design antibody-toxin complexes to  
 CC target tumour cells in these tissues. The AMH-receptor sequences can be  
 CC used to screen and purify compounds that bind to them, thereby  
 CC identifying AMH ligands. The antibody designed from the receptor  
 CC sequences can be used in immunoassays to detect the levels of  
 CC AMH-receptor. The antibody can also be linked to a toxin in order to  
 CC kill cells that express the receptor, e.g. cancer cells. The advantage  
 CC of using antibody conjugates based on AMH-receptors are that they can be  
 CC used more aggressively than those based on antibodies that recognise  
 CC receptors more generally distributed.  
 SQ Sequence 2228 BP; 393 A; 745 C; 646 G; 444 T;

Query Match 2.3%; Score 18; DB 1; Length 2228;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 gctggtcctgtgtgct 130  
 |||||  
 Db 337 GCTGGTCCTGCTGCT 354

## RESULT 7

Q98987  
 ID Q98987 standard; cDNA: 2408 BP.  
 AC Q98987;  
 DT 27-FEB-1996 (first entry)  
 DE H1 AMH-receptor cDNA.  
 KW 2B10 anti-Mullerian hormone receptor; anti-Mullerian hormone; AMH;  
 KW receptor; antibody; therapy; H1; 2B10; tumour;  
 KW transforming growth factor-beta; TGF-beta; bone morphogenic protein;  
 KW activin; Mullerian inhibiting substance; MIS; ss.  
 OS Oryctolagus cuniculus.  
 FH Key Location/Qualifiers  
 FT cds 59..1768  
 FT /tag- a  
 FT /product- H1 AMH receptor  
 FT signal\_peptide 59..109  
 FT /tag- b  
 FT mat\_peptide 110..1765  
 FT /tag- c  
 FT /product- AMH-receptor  
 FT misc\_feature 110..1765  
 FT /tag- d  
 FT /note- "claimed"  
 FT misc\_feature 110..490  
 FT /tag- e  
 FT /note- "extracellular domain"  
 FT misc\_feature 491..568  
 FT /tag- f  
 FT /note- "transmembrane region"  
 FT misc\_feature 569..1765  
 FT /tag- g  
 FT /note- "cytoplasmic domain"  
 FT polya\_signal 2390..2395

```

FT WO9516709-A2. /*tag= h
PN 22-JUN-1995.
PD 13-DEC-1994; U14643.
PF 13-DEC-1993; US-166333.
PR 23-DEC-1993; US-173512.
PA (BIOJ ) BIOGEN INC.
PA (INRM ) INSERM INST NAT SANTE & RECH MED.
PI Cate RL, Jesso N;
DR WPI; 95-231521/30.
DR P-PSDB; R80632.
DR New DNA encoding anti-Mullerian hormone receptor, related
PT polypeptide(s) and antibodies - useful in immunoassays screening for
PT ligands, treatment of cancer cells expressing the receptor etc.
PS Claim 1; Page 43-44; 77pp; English.
CC This sequence represents the cDNA encoding the H1 anti-Mullerian hormone
CC (AMH) receptor isoform. The difference between this isoform and the 2B10
CC isoform is that this sequence contains an extra exon encoding 61 amino
CC acids. AMH is a glycoprotein and is part of the transforming growth
CC factor-beta superfamily. AMH is structurally and functionally related to
CC TGF-beta bone morphogenic protein, and activin. AMH receptors are
CC present in a limited number of tissues and can therefore be used to
CC design antibody-toxin complexes to target tumour cells in these tissues.
CC The AMH-receptor sequences can be used to screen and purify compounds
CC that bind to them, thereby identifying AMH ligands. The antibody
CC designed from the receptor sequences can be used in immunoassays to
CC detect the levels of AMH-receptor. The antibody can also be linked to a
CC toxin in order to kill cells that express the receptor, e.g. cancer
CC cells. The advantage of using antibody conjugates based on AMH-receptors
CC are that they can be used more aggressively than those based on
CC antibodies that recognise receptors more generally distributed.
SQ Sequence 2408 BP; 431 A; 798 C; 705 G; 474 T;

Query Match 2.3%; Score 18; DB 1; Length 2408;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 gctggtctgctgctgt 130
Db 517 GCTGGCTCTGCTGCTGCT 534

RESULT 8
T11342/c
ID T11342 standard; DNA; 3663 BP.
AC T11342;
DE 08-JAN-1997 (first entry)
KW nod gene fragment of Rhizobium trifolii.
KW Legume exudate-inducible promoter; promoter; nod; Rhizobium; toxin;
KW insecticide; Bacillus thuringiensis; gene expression; recombinant;
KW hydrogenase; metallothionein; prolactin; ds.
OS Rhizobium trifolii.
FH Key Location/Qualifiers
FT cds complement (1. .336)
FT /*tag= a
FT /product= Nod C protein.
FT complement (360. .1004)
FT /*tag= b
FT /product= NodB protein
FT complement (1003. .1591)
FT /*tag= c
FT /product= NodA protein
FT 1826. .2779
FT /*tag= d
FT /product= NodD protein
FT 3265. .3541
FT /*tag= e
FT /product= NodF protein
FT note= "3544"
FT 3544. .3663
FT /*tag= f
FT /product= NodE protein

```

```

PN US5484718-A.
PD 16-JAN-1996.
PF 17-JUN-1986; 875300.
PR 17-JUN-1986; US-875300.
PA (MYCO ) MYCOGEN PLANT SCI INC.
PI Djordjevic MA, Innes RW, Kuempel PL, Rolfe BG, Schofield PR;
PI Scott KE, Watson JM;
DR WPI; 96-087064/09.
DR P-PSDB; R98751, R98160, R98161, R98209, R98210, W06150.
DR Gene expression using legume exudate-inducible promoter - useful for
PT expressing proteins, esp. Bacillus thuringiensis toxin, on contact
PT with legume exudate.
PS Disclosure: Figure 2: 32pp; English.
CC Expressing foreign structural genes using a legume exudate-inducible
CC promoter comprises placing the structural gene under the control of
CC such a promoter and inserting the construct into a recombinant DNA
CC molecule comprising a nod D gene of a strain of Rhizobium. The
CC construct is then introduced into a bacterial strain in which the
CC promoter is active and the nod D gene is expressed. The bacterial
CC strain is then combined with an effective amount of a legume
CC gene inducing composition effective for induction of a legume
CC exudate-inducible gene. The promoter is preferably the promoter of
CC the nod ABC or nod FE genes of Rhizobium trifoli. The structural
CC gene is preferably an insect toxin gene of Bacillus thuringiensis.
CC This sequence is the promoter region between nodD and the nodKABC
CC gene cluster in Bradyrhizobium sp. (Parasponia) ANU 289. Examples
CC of other proteins that can be usefully expressed using such methods
CC include: hydrogenase, metallothionein and prolactin.
SQ Sequence 3663 BP; 832 A; 1019 C; 994 G; 818 T;

Query Match 2.3%; Score 18; DB 1; Length 3663;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 gtggccttctggatcatt 494
Db 1046 GTGGCCTTCGGATCAT 1029

RESULT 9
X29464
ID X29464 standard; DNA; 3736 BP.
AC X29464;
DE 10-JUN-1999 (first entry)
DE Human negative regulatory factor (NRF) encoding DNA.
DE Transcription factor protein; negative regulatory factor; NRF; NRE;
KW NF-kB; negative regulatory element; rheumatoid arthritis; inflammatory;
KW ischemia; hemorrhagic; septic shock; allograft rejection; meningitis;
KW infection; pulmonary; tumour; genetic disease; gene therapy; human; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 654. .1820
FT /*tag= a
FT misc_feature 984. .1077
FT /*tag= b
FT /note= "It is stated in the claims that this
FT fragment can be replaced by a fragment
FT from murine NRF"
FT misc_feature 1897. .1979
FT /*tag= c
FT /note= "It is stated in the claims that this
FT fragment can be replaced by a fragment
FT from murine NRF"
FT misc_feature 654. .1526
FT /*tag= d
FT /note= "fragment coding for silencer domain; claimed in
FT claim 2"
FT misc_feature 654. .1817
FT /*tag= e
FT /note= "region claimed in claim 2"
FT misc_feature 1518. .1817
FT /*tag= f

```

```

FT /note= "fragment coding for DNA binding domain; claimed
FT in claim 2"
PD
PN EP-094853-A1.
PD 03-FEB-1999.
PP 24-JUL-1997; 112704.
PP 24-JUL-1997; EP-112704.
PA (GBPB ) GBS BIOTECHNOLOGISCHE FORSCHUNG MBH.
PI Hauser H, Nourbakhsh M;
DR WPI; 99-108345/10.
DR P-PSDB; Y03753.
DR New inhibitory transcription factor NRF modulates NF-kB family
PT members - useful for treatment of inflammatory disorders, rheumatoid
PT arthritis, tumours and genetic diseases
PS Claim 1; Fig 1B; 30pp; English.
CC This DNA encodes a transcription factor protein NRF (negative regulatory
CC factor). The NRF silences transcriptional activity and modulates NF-kB
CC family members by binding negative regulatory element (NRE)-motifs in
CC human gene promoters. The NRF DNA, protein and vectors comprising the DNA
CC are useful for identifying and developing agonists and antagonists of
CC NRF-functions, and for developing improved antisense NRF and ribozymes.
CC They are also useful for the detection and diagnosis of transient or
CC permanent regulatory disorders of NF-kB-related and/or NRF-regulated
CC physiological patterns in animals or humans, and for therapy development
CC and treatment of diseases, especially rheumatoid arthritis, inflammatory
CC disorders including ischemia, hemorrhagic and septic shock, allograft
CC rejection, bacterial meningitis, acute airway inflammation, infectious
CC diseases, pulmonary complications induced by cardiopulmonary bypass,
CC tumours and genetic diseases using gene therapy with vectors to transform
CC cells or organisms for transient or permanent expression of the protein.
SQ Sequence 3736 BP; 1212 A; 684 C; 754 G; 1086 T;

Query Match 2.3%; Score 18; DB 1; Length 3736;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 748 accctgccccaagcacca 765
Db 1066 ACCCTGCCCAAGCACCA 1083
|||||

RESULT 10
Q43661
ID Q43661 standard; DNA; 4558 BP.
AC Q43661;
DT 13-OCT-1993 (first entry)
DE Acetobacter cdg2 operon.
KW Cyclic diguanylate; diguanylate phosphodiesterase;
KW diguanylate cyclase; cellulose production; ss.
OS Acetobacter xylinum.
PH Key Location/Qualifiers
FT cds 98..2354
FT /tag= a
FT /label= pdeA2
FT /note= "a diguanylate phosphodiesterase gene;
FT the amino acid sequence R38153 cannot be
FT deduced directly from this open reading
FT frame due to a number of apparent
FT (typographical ?) insertions and deletions
FT in the printed sequence"
FT misc_difference 326..327
FT /tag= b
FT /codon= seq: GG; aa: Val
FT /note= "inserting a T between nucleotides 326 and
FT 327 will restore the reading frame"
FT misc_difference 730..732
FT /tag= c
FT /codon= seq: TAG; aa: Met
FT /note= "inserting a T between nucleotides 730 and
FT 732 will restore the reading frame"
FT misc_difference 739..740
FT /tag= d
FT /codon= seq: GC; aa: Arg
FT /note= "inserting a C between nucleotides 738 and
FT 739 will restore the reading frame"

```

```

FT misc_difference 789..791
FT /tag= e
FT /codon= seq: ATC; aa: Met
FT misc_difference 792..794
FT /tag= f
FT /codon= seq: GTG; aa: Leu
FT misc_difference 834..835
FT /tag= g
FT /codon= seq: GC; aa: Gly
FT /note= "inserting a G between nucleotides 834 and
FT 835 will restore the reading frame"
FT misc_difference 908
FT /tag= h
FT /codon= seq: CAGG; aa: Gln
FT /note= "delete G(908) to maintain reading frame"
FT misc_difference 918..921
FT /tag= i
FT /codon= seq: GCTG; aa: Val
FT /note= "delete C(919) to maintain reading frame"
FT misc_difference 1258..1260
FT /tag= j
FT /codon= seq: GAA; aa: Asp
FT misc_difference 1720..1721
FT /tag= k
FT /codon= seq: CC; aa: Ala
FT /note= "inserting a G between nucleotides 1719 and
FT 1720 will restore reading frame"
FT cds 2463..4187
FT /tag= l
FT /label= dgc2
FT /note= "a diguanylate cyclase gene"
FT
FT W09311244-A.
PD 10-JUN-1993.
PD 14-OCT-1992; U08756.
PP 29-NOV-1991; US-800218.
PA (WEYE ) WEYERHAEUSER CO.
PI Ben-Bassat A, Benziman M, Calhoun RD, Gelfand DH;
PI Tal R, Wong HC;
DR WPI; 93-197062/24.
DR P-PSDB; R38153, R38154.
DR Polynucleotide sequence from Acetobacter cdg operon - encodes
FT cyclic di:guanosine mono:phosphate degradation enzymes e.g.
FT 3-phosphodiesterase isozyme
PS Claim 3; Page 67-70; 98pp; English.
CC A gene bank of Acetobacter 1306-11 was constructed in the broad host
CC range, mobilisable cosmid pKT230C085. The average insert size was ca.
CC 30kb. The bank was screened with a 53bp PCR amplified probe whose
CC sequence was expected to be an exact match for a diguanylate
CC cyclase encoding gene. Three unique cosmid classes were identified
CC corresponding to three different cdg operons. The cdg2 operon
CC comprises at least 2 genes. Manipulating the expression levels of
CC the different cdg2 genes is contemplated, particularly in cellulose
CC producing bacteria.
SQ Sequence 4558 BP; 850 A; 1379 C; 1467 G; 862 T;

Query Match 2.3%; Score 18; DB 1; Length 4558;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 562 accgcctgcaggccatcc 579
|||||
Db 3196 ACCGCCGACAGGCCATCC 3213

RESULT 11
T89253/C
ID T89253 standard; cDNA; 5294 BP.
AC T89253;
DT 27-MAR-1998 (first entry)
DE cDNA for human CASPR/p190, a ligand for RPTP-beta.
KW CASPR; p190; contactin associated protein; human;
KW signal transduction; transmembrane protein tyrosine phosphatase;

```

KW RPT-beta; neurite; neuron; neurological disease;  
KW Parkinson's disease; Alzheimer's disease Huntington's disease;  
KW cerebral palsy; mental retardation; amyotrophic lateral sclerosis;  
KW inflammation; arthritis; colorectal cancer; therapy; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 218..4372  
FT /tag= a  
FT sig\_peptide 218..274  
FT /tag= b  
FT mat\_peptide 274..4369  
FT /tag= c  
PN WO9735872-A1.  
PD 02-OCT-1997.  
PF 27-MAR-1997; U05270.  
PR 26-MAR-1997; US-826134.  
PR 27-MAR-1996; US-014199.  
PA (SUGEN-) SUGEN INC.  
PI Peles E;  
DR WPI; 97-489558/45.  
DR P-PSDB; W31946.  
PT Nucleic acid encoding human and rat p190 and related vectors -  
PT involved in signalling that controls neurite growth, differentiation  
PT and survival in neurons, for treatment of neurological disease  
PS Claim 1, Page 80-89; 118pp; English.  
CC This cDNA clone codes for human 190 kDa contactin associated  
CC protein, or CASPR/p190 (see W31946), a protein implicated as the  
CC bridge between contactin and intracellular second messenger systems  
CC for the signal caused by the binding of the carboxy anhydride  
CC domain of RPT-beta (a transmembrane protein tyrosine phosphatase)  
CC to contactin and resulting in neurite growth, differentiation or  
CC survival. The clone was isolated from human IMR32 neuroblastoma  
CC cells using probes from the rat CASPR/p190 sequence (see T89254).  
CC CASPR/p190 nucleic acids, including those that encode amino acids  
CC 40-168, 199-330, 362-486, 544-576, 582-739, 809-938, 961-985,  
CC 1031-1077, 1083-1218, 1282-1306 and 1328-1369 of human CASPR/p190,  
CC can be used to produce recombinant proteins in eukaryotic or  
CC prokaryotic host cells. p190, or its mimics, enhancers and  
CC suppressors, are used to treat neurological disease, or p190  
CC expression is controlled using gene therapy, e.g. expression of  
CC ribozymes or antisense molecules. Conditions that can be treated  
CC include Parkinson's, Alzheimer's and Huntington's diseases,  
CC cerebral palsy, mental retardation and amyotrophic lateral  
CC sclerosis. p190 can also be used to treat inflammation, e.g. the  
CC arthritis, colorectal cancer, and vascular inflammation, e.g. the  
CC result of arteriosclerosis, angioplasty, myocardial infarction etc.  
CC p190 is administered, e.g. by injection, to the treatment site, or  
CC when nucleic acid or vectors are used, these are attached to a  
CC regulatory sequence that directs tissue-specific expression. Drugs  
CC that modify growth, differentiation and survival of neurons are  
CC assessed from their activity on p190, using cells or microsomal  
CC extracts. Also 'pathway genes' associated with p190 gene products  
CC can be identified, e.g. in a standard yeast two-hybrid assay.  
SQ Sequence 5294 BP; 1099 A; 1576 C; 1446 G; 1173 T;

Query Match 2.3%; Score 18; DB 1; Length 5294;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 701 gctgtagggtgggacc 718  
Db 4257 GCTGTAGGCTGGGACC 4240

## RESULT 12

ID T40262 standard; CDNA; 39 BP.  
AC T40262;

DE 20-NOV-1996 (first entry)  
KW Human ob protein forward primer OB.F2H.

KW Human; obesity protein; ob; fat cell cDNA library; primer; PCR;  
KW polymerase chain reaction; amplify; dipeptide leader; antibody;

KW peripheral adipose tissue; weight gain; obese; peptide hormone;  
KW fat deposit; type II diabetes; cardiovascular disease; cancer; ss.  
OS Synthetic.  
PN WO9624670-A1.  
PD 15-AUG-1996.  
PF 29-JAN-1996; U01411.  
PR 06-FEB-1995; US-384493.  
PR 26-MAY-1995; US-451250.  
PA (ELIL ) LILLY & CO ELI.  
PI Basinski MB, Schoner BE;  
DR WPI; 96-384442/38.  
PT New DNA encoding two forms of the human obesity gene product - and  
PT related vectors and transformed cells, used to produce recombinant  
PT proteins for treating obesity  
PS Example 1; Page 17; 30pp; English.  
CC The sequences given in T40261-64 are primers which amplify the ob  
CC protein coding sequences given in T40257-60. T40257-58 encode a  
CC novel form of the human obesity protein, ob and T40259-60 encode  
CC another novel form of the ob protein in which Ala27Gln28 are replaced  
CC by a single Thr. These nucleic acid sequences were amplified from  
CC a human fat cell cDNA library. The ob proteins are pref. expressed  
CC with a dipeptide leader sequence of Met-Arg or Met-Tyr. Experiments  
CC have suggested that the ob proteins are released by peripheral  
CC adipose tissue and are capable of controlling weight gain in normal  
CC and obese mice. The ob proteins are thought to be circulating peptide  
CC hormones which regulate the size of a bodies fat deposits. They can  
CC be used to treat obesity and to reduce the risk of type II diabetes,  
CC cardiovascular disease and cancer. Anti-ob protein antibodies may be  
CC used for diagnostic purposes.  
SQ Sequence 39 BP; 11 A; 8 C; 15 G; 5 T;

Query Match 2.2%; Score 17; DB 1; Length 39;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 425 ggtaccatccagaag 441  
Db 13 GGTACCCATCCAGAGG 29

## RESULT 13

ID T51178 standard; DNA; 39 BP.  
AC T51178;

DE 16-OCT-1997 (first entry)  
KW Human leptin PCR forward primer OB.F2H.

KW Type II diabetes; endogenous circulating leptin; obesity gene;  
KW porcine; bovine; human; rhesus; rat; leptin mimetic; murine;  
KW polymerase chain reaction; ss.

OS Synthetic.  
PN WO9702004-A2.

PD 23-JAN-1997.  
PF 26-JUN-1996; U10983.

PR 30-JUN-1995; US-000718.  
PA (ELIL ) LILLY & CO ELI.

PI Stephens TW;  
DR WPI; 97-108725/10.

PT Use of leptin for treatment or prevention of diabetes mellitus -  
PT esp. type II diabetes associated with low levels of endogenous  
PT circulating leptin

PS Example 1; Page 21; 33pp; English.

CC The present sequence represents the forward PCR primer OB.F2H for  
CC the amplification of human leptin, which is a preferred leptin for  
CC use in a method of treating or preventing diabetes mellitus. The  
CC method involves the administering of leptin, or its mimetics, to  
CC patients with type II diabetes, associated with high or particularly  
CC low levels of circulating endogenous leptin, preferably 0-80,  
CC especially 0-15, ng/ml.  
SQ Sequence 39 BP; 11 A; 8 C; 15 G; 5 T;

Query Match 2.2%; Score 17; DB 1; Length 39;

DK WFL; 30-11/04//12.

DK WFL; 30-11/04//12.

Job time: 18921 sec

Job time: 18921 sec

Job time: 18921 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:10:59 ; Search time 181.17 Seconds  
(without alignments)  
518.754 Million cell updates/sec

Title: US-09-215-435-118

Perfect score: 785

Sequence: 1 cgggaatccgggagtcgggtg.....tatggaaataaagttcttc 785

Scoring table:

OLIGO\_NUC

Gapop 60.0 , Capext 60.0

Searched: 214294 seqs, 59861574 residues

Word size : 0

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued Patents, NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PCFUS9\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	18	2.3	1033	3	US-08-807-050-2
c 2	18	2.3	4558	2	US-08-309-512-2
c 3	18	2.3	4558	6	PCT-US92-08756A-2
c 4	17	2.2	39	2	US-08-678-369-10
c 5	17	2.2	39	4	US-09-047-243-10
c 6	17	2.2	1149	1	US-08-474-177-3
c 7	17	2.2	1149	2	US-08-487-033-3
c 8	17	2.2	1149	2	US-08-480-810-3
c 9	17	2.2	1149	3	US-08-508-735-3
c 10	17	2.2	1149	3	US-08-848-251-3
c 11	17	2.2	1149	4	US-08-486-047-3
c 12	17	2.2	1785	4	US-08-910-856-1
c 13	17	2.2	1787	4	US-08-808-982-2
c 14	17	2.2	2012	4	US-08-555-568B-16
c 15	17	2.2	2109	4	US-08-555-568B-20
c 16	17	2.2	2112	4	US-08-555-568B-22
c 17	17	2.2	3187	4	US-08-910-856-9
c 18	17	2.2	3187	4	US-08-910-856-10
c 19	17	2.2	6192	3	US-08-479-537A-1
c 20	17	2.2	6449	3	US-08-479-537A-4
c 21	17	2.2	246240	3	US-08-724-394A-20
c 22	17	2.2	246240	3	US-08-724-394A-21
c 23	17	2.2	246240	3	US-08-724-394A-22
c 24	16	2.0	431	5	US-09-026-343-26
c 25	16	2.0	455	5	US-09-026-343-20
c 26	16	2.0	487	5	US-09-026-343-18

c 27	16	2.0	615	1	US-08-081-072-2	Sequence 2, Appli
c 28	16	2.0	615	1	US-08-449-093A-2	Sequence 2, Appli
c 29	16	2.0	838	1	US-07-590-894C-1	Sequence 1, Appli
c 30	16	2.0	915	7	5304637-2	Patent No. 5304637
c 31	16	2.0	1035	4	US-08-993-228-32	Sequence 32, Appli
c 32	16	2.0	1426	1	US-08-081-072-5	Sequence 5, Appli
c 33	16	2.0	1426	1	US-08-449-093A-5	Sequence 5, Appli
c 34	16	2.0	2046	3	US-08-911-445-6	Sequence 6, Appli
c 35	16	2.0	2075	1	US-08-167-628-1	Sequence 1, Appli
c 36	16	2.0	2075	1	US-08-386-680-1	Sequence 1, Appli
c 37	16	2.0	2075	2	US-08-459-717-1	Sequence 1, Appli
c 38	16	2.0	2075	2	US-08-712-302-1	Sequence 1, Appli
c 39	16	2.0	2075	4	US-08-880-031-1	Sequence 1, Appli
c 40	16	2.0	2075	6	PCT-US96-08140-1	Sequence 1, Appli
c 41	16	2.0	2200	3	US-08-819-825-1	Sequence 1, Appli
c 42	16	2.0	2797	4	US-09-021-323-2	Sequence 2, Appli
c 43	16	2.0	3134	3	US-08-533-669A-1	Sequence 1, Appli
c 44	16	2.0	3393	1	US-08-295-502-1	Sequence 1, Appli
c 45	16	2.0	3393	6	PCT-US95-10691-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-08-807-050-2/c  
; Sequence 2, Application US/08807050  
; Patent No. 5869256  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: HUMAN COATOMER VESICLE PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/807,050  
; FILING DATE: Herewith  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0226 US  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1033 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: COLNF02  
; CLONE: 1311434  
; US-08-807-050-2

Query Match 2.3% Score 18; DB 3; Length 1033;

Best Local Similarity 100.0%; Pred. No. 7.7; Mismatches 0; Indels 0; Gaps 0;

QY 642 cactccaggctgtccccc 659  
 Db 456 CACTCCAGGCTGTCCCCC 439

RESULT 2

US-08-309-512-2  
 ; Sequence 2, Application US/08309512  
 ; Patent No. 5759828  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tal, Ronny  
 ; APPLICANT: Benziman, Moshe  
 ; APPLICANT: Gelfand, David H.  
 ; APPLICANT: Ben-Bassat, Arie  
 ; APPLICANT: Calhoon, Roger D.  
 ; APPLICANT: Wong, Hing C.  
 ; TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES  
 ; NUMBER OF SEQUENCES: 63  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennie & Edmonds  
 ; STREET: 2730 Sand Hill Road  
 ; CITY: Menlo Park  
 ; STATE: California  
 ; COUNTRY: U.S.A.  
 ; ZIP: 94025  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/309,512  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/800,218  
 ; FILING DATE: 29-NOV-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Bortner, Scott R.  
 ; REGISTRATION NUMBER: 34,298  
 ; REFERENCE/DOCKET NUMBER: 8145-008  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 854-3660  
 ; TELEFAX: (415) 854-3694  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4558 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Acetobacter xylinum  
 ; US-08-309-512-2

Query Match 2.3%; Score 18; DB 2; Length 4558;  
 Best Local Similarity 100.0%; Pred. No. 7.2; Mismatches 0; Indels 0; Gaps 0;

QY 562 accgcctgcaggccatcc 579  
 Db 3196 ACCGCCTGCAGGCCATCC 3213

RESULT 3

PCT-US92-08756A-2

; Sequence 2, Application PC/TUS9208756A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tal, Ronny  
 ; APPLICANT: Benziman, Moshe  
 ; APPLICANT: Gelfand, David H.  
 ; APPLICANT: Ben-Bassat, Arie  
 ; APPLICANT: Calhoon, Roger D.  
 ; APPLICANT: Wong, Hing C.  
 ; TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Limbach and Limbach  
 ; STREET: 2001 Ferry Building  
 ; CITY: San Francisco  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94111  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0,  
 ; SOFTWARE: Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US92/08756A  
 ; FILING DATE: 19921014  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/800,218  
 ; FILING DATE: 29-NOV-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Bortner, Scott R.  
 ; REGISTRATION NUMBER: 34,298  
 ; REFERENCE/DOCKET NUMBER: WEYR 20050 USA  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-433-4150  
 ; TELEFAX: 415-433-8716  
 ; TELEX: 278356  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4558 base pairs  
 ; TYPE: NUCLEIC ACID  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Acetobacter xylinum  
 ; PCT-US92-08756A-2

Query Match 2.3%; Score 18; DB 6; Length 4558;  
 Best Local Similarity 100.0%; Pred. No. 7.2; Mismatches 0; Indels 0; Gaps 0;

QY 562 accgcctgcaggccatcc 579  
 Db 3196 ACCGCCTGCAGGCCATCC 3213

RESULT 4

US-08-678-369-10  
 ; Sequence 10, Application US/08678369  
 ; Patent No. 5756461  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Stephens, Thomas W.  
 ; TITLE OF INVENTION: METHODS FOR TREATING DIABETES  
 ; NUMBER OF SEQUENCES: 12  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Eli Lilly and Company  
 ; STREET: Lilly Corporate Center/Patent Division

```

; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/678,369
; FILING DATE: 28-JUN-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Maciak M., Ronald S.
; REGISTRATION NUMBER: 35,262
; REFERENCE/DOCKET NUMBER: X-10289
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-1664
; TELEFAX: 317-277-1917
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-678-369-10
;
; Query Match 2.2%; Score 17; DB 2; Length 39;
; Best Local Similarity 100.0%; Pred. No. 27;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 425 ggtaccatccagaag 441
; Db 13 GGTACCATCCAGAGG 29
;
; RESULT 5
; US-09-243-10
; Sequence 10, Application US/09047243
; Patent No. 5922678
; GENERAL INFORMATION:
; APPLICANT: Stephens, Thomas W.
; TITLE OF INVENTION: METHODS FOR TREATING DIABETES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center/Patent Division
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/047,243
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/678,369
; FILING DATE: 28-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Maciak M., Ronald S.
; REGISTRATION NUMBER: 35,262
; REFERENCE/DOCKET NUMBER: X-10289
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-1664
; TELEFAX: 317-277-1917

```

```

; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-09-047-243-10
;
; Query Match 2.2%; Score 17; DB 4; Length 39;
; Best Local Similarity 100.0%; Pred. No. 27;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 425 ggtaccatccagaag 441
; Db 13 GGTACCATCCAGAGG 29
;
; RESULT 6
; US-08-474-177-3/c
; Sequence 3, Application US/08474177
; Patent No. 5624819
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H. Lisa A.
; APPLICANT: Cannon-Albright, Lisa A.
; TITLE OF INVENTION: GERMLINE MUTATIONS IN THE MTS GENE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,177
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/03537
; FILING DATE: 17-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/251,938
; FILING DATE: 01-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/215,087
; FILING DATE: 18-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/215,086
; FILING DATE: 18-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,369
; FILING DATE: 14-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/214,582
; FILING DATE: 18-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109348-E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1149 base pairs

```

;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; ORIGINAL SOURCE:  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: 5'UTR  
;; LOCATION: 1..866  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 867..1016  
;; FEATURE:  
;; NAME/KEY: intron  
;; LOCATION: 1017..1149  
US-08-474-177-3

Query Match 2.2%; Score 17; DB 1; Length 1149;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 cctcccacctgcccc 98  
|||||  
Db 524 CTTCCCACTGCCCC 508

RESULT 7  
US-08-487-033-3/c  
; Sequence 3, Application US/08487033  
; Patent No. 5739027  
; GENERAL INFORMATION:  
; APPLICANT: Ramb, Alexander  
; TITLE OF INVENTION: MTS1-Beta GENE  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
; STREET: 1201 New York Avenue, Suite 1000  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487,033  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/03316  
; FILING DATE: 17-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/251,938  
; FILING DATE: 01-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/215,087  
; FILING DATE: 18-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/215,086  
; FILING DATE: 18-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/227,369  
; FILING DATE: 14-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/214,582  
; FILING DATE: 18-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.

;; REGISTRATION NUMBER: 28,957  
;; REFERENCE/DOCKET NUMBER: 24884-109348-C  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-962-4810  
;; TELEFAX: 202-962-8300  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1149 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; ORIGINAL SOURCE:  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: 5'UTR  
;; LOCATION: 1..866  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 867..1016  
;; FEATURE:  
;; NAME/KEY: intron  
;; LOCATION: 1017..1149  
US-08-487-033-3

Query Match 2.2%; Score 17; DB 2; Length 1149;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 cctcccacctgcccc 98  
|||||  
Db 524 CTTCCCACTGCCCC 508

RESULT 8  
US-08-480-810-3/c  
; Sequence 3, Application US/08480810  
; Patent No. 5801236  
; GENERAL INFORMATION:  
; APPLICANT: Ramb, Alexander  
; TITLE OF INVENTION: MTS1 GENE  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
; STREET: 1201 New York Avenue, Suite 1000  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,810  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/03316  
; FILING DATE: 17-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/251,938  
; FILING DATE: 01-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/215,087  
; FILING DATE: 18-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/215,086  
; FILING DATE: 18-MAR-1994

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,369
; FILING DATE: 14-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/214,582
; FILING DATE: 18-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109348
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-8300
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1149 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..866
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 867..1016
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1017..1149
;
US-08-480-810-3
```

```
Query Match 2.2%; Score 17; DB 2; Length 1149;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 cctcccccacctgcccc 98
| | | | | | | | | | | | | | | |
Db 524 CCTCCCCACCTGCCCC 508
```

```
RESULT 9
US-08-735-3/c
; Sequence 3, Application US/08508735
; Patent No. 5843756
; GENERAL INFORMATION:
; APPLICANT: Stone, Steven
; APPLICANT: Jiang, Ping
; APPLICANT: Kamb, Alexander
; TITLE OF INVENTION: MTS GENE AND THERAPEUTIC USE THEREOF
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/508,735
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US to be assigned
```

```
;
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/03316
; FILING DATE: 17-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109348
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4848
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1149 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..866
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 867..1016
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1017..1149
;
US-08-508-735-3
```

```
Query Match 2.2%; Score 17; DB 3; Length 1149;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 cctcccccacctgcccc 98
| | | | | | | | | | | | | | | |
Db 524 CCTCCCCACCTGCCCC 508
```

```
RESULT 10
US-08-848-251-3/c
; Sequence 3, Application US/08848251
; Patent No. 5989815
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Cannon-Albright, Lisa A.
; APPLICANT: Kamb, Alexander
; TITLE OF INVENTION: GERMLINE MUTATIONS IN THE MTS GENE AND
; TITLE OF INVENTION: METHOD FOR DETECTING PREDISPOSITION TO CANCER AT THE MTS
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/848,251
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/474,083
```

```

; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: PCT/US95/03537
; FILING DATE: 17-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/251,938
; FILING DATE: 01-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/215,087
; FILING DATE: 18-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/215,086
; FILING DATE: 18-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,369
; FILING DATE: 14-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/214,582
; FILING DATE: 18-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109348-G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1149 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..866
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 867..1016
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1017..1149
; US-08-848-251-3

```

```

Query Match 2.2%; Score 17; DB 4; Length 1149;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 82 cctcccacactgcccc 98
DB 524 CCTCCCCACCTGCCCC 508

```

```

RESULT 11
US-08-486-047-3/c
; Sequence 3, Application US/08486047
; Patent No. 5994095
; GENERAL INFORMATION:
; APPLICANT: Kamb, Alexander
; TITLE OF INVENTION: WTS2 GENE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,047
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/03316
; FILING DATE: 17-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/251,938
; FILING DATE: 01-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/215,087
; FILING DATE: 18-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/215,086
; FILING DATE: 18-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,369
; FILING DATE: 14-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/214,582
; FILING DATE: 18-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109348-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1149 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..866
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 867..1016
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1017..1149
; US-08-486-047-3

```

```

Query Match 2.2%; Score 17; DB 4; Length 1149;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 82 cctcccacactgcccc 98
DB 524 CCTCCCCACCTGCCCC 508

```

```

RESULT 12
US-08-910-856-1
; Sequence 1, Application US/08910856
; Patent No. 5981257
; GENERAL INFORMATION:
; APPLICANT: FUKUI, TOSHIKI
; APPLICANT: DOI, YOSHIHARU
; TITLE OF INVENTION: POLYESTER SYNTHASE GENE AND PROCESS

```



;; TITLE OF INVENTION: FOR PRODUCING POLYESTER  
;; NUMBER OF SEQUENCES: 16  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: FISH & RICHARDSON P.C.  
;; STREET: 4225 EXECUTIVE SQUARE, SUITE 1400  
;; CITY: LA JOLLA  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 92037  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/910,856  
;; FILING DATE: 13-AUG-1997  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 214509/1996  
;; FILING DATE: 14-AUG-1996  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 199979/1997  
;; FILING DATE: 25-JUL-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: HALL, Lisa A.  
;; REGISTRATION NUMBER: 38,347  
;; REFERENCE/DOCKET NUMBER: 07898/016001  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 619-678-5070  
;; TELEFAX: 619-678-5099  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1785 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..1782  
US-08-910-856-1

Query Match 2.2%; Score 17; DB 4; Length 1785;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 521 ggatgcctgagggcg 537  
Db 378 GGATGCCCTGGAGGCG 394

RESULT 13  
US-08-808-982-2  
; Sequence 2, Application US/08808982  
; Patent No. 5939271  
; GENERAL INFORMATION:  
; APPLICANT: Tessier-Lavigne, Marc  
; APPLICANT: Leonardo, E. David  
; APPLICANT: Hink, Lindsay  
; APPLICANT: Masu, Masayuki  
; APPLICANT: Kazuko, Keino-Masu  
; TITLE OF INVENTION: Netrin Receptors  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/808,982  
;; FILING DATE:  
;; CLASSIFICATION: 530  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: OSMAN, RICHARD A.  
;; REGISTRATION NUMBER: 36,627  
;; REFERENCE/DOCKET NUMBER: UC96-217  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 343-4341  
;; TELEFAX: (415) 343-4342  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1787 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: CDNA  
US-08-808-982-2

Query Match 2.2%; Score 17; DB 4; Length 1787;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ctggctctgctgctgct 130  
Db 89 CTGGCTCTGCTGCTGCT 105

RESULT 14  
US-08-555-568B-16  
; Sequence 16, Application US/085555568B  
; Patent No. 5976854  
; GENERAL INFORMATION:  
; APPLICANT: Jones, Simon  
; APPLICANT: Tang, Jim  
; TITLE OF INVENTION: Calcium Independent Phospholipase A2/B  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/555,568B  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brown, Scott A.  
; REGISTRATION NUMBER: 32,724  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 498-8224  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2012 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHETICAL: NO

```

;
; NAME/KEY: CDS
; LOCATION: 43..1224
US-08-555-568B-16

Query Match      2.2%; Score 17; DB 4; Length 2102;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 gcctccccacctgcccc 97
    | | | | | | | | | | |
DB 19 GCCTCCCACTGCCCC 35

```

```

RESULT 15
US-08-555-568B-20
; Sequence 20, Application US/08555568B
; Patent No. 5976854
; GENERAL INFORMATION:
; APPLICANT: Jones, Simon
; APPLICANT: Tang, Jim
; TITLE OF INVENTION: Calcium Independent Phospholipase A2/B
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA: US/08/555,568B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2109 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 43...2103
US-08-555-568B-20

```

```

Query Match      2.2%; Score 17; DB 4; Length 2109;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 gcctccccacctgcccc 97
    | | | | | | | | | | |
DB 19 GCCTCCCACTGCCCC 35

```

Search completed: May 1, 2000, 16:11:07  
Job time: 18732 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 20:00:08 ; Search time 4088.29 Seconds  
(without alignments)  
-658.450 Million cell updates/sec

Title: US-09-215-435-118  
Perfect score: 785  
Sequence: 1 cggatccggagtcggtg.....tatggaataaagttcttc 785

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 5265378 seqs, -1714608768 residues

Word size : 0

Total number of hits satisfying chosen parameters: 10530736

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Pending\_Patents\_NA:\*

1: /cgn2\_6/ptodata/1/pna/PCTUS9\_COMB.seq:\*

2: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq:\*

3: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq:\*

4: /cgn2\_6/ptodata/1/pna/US081A\_COMB.seq:\*

5: /cgn2\_6/ptodata/1/pna/US081B\_COMB.seq:\*

6: /cgn2\_6/ptodata/1/pna/US081C\_COMB.seq:\*

7: /cgn2\_6/ptodata/1/pna/US081D\_COMB.seq:\*

8: /cgn2\_6/ptodata/1/pna/US082A\_COMB.seq:\*

9: /cgn2\_6/ptodata/1/pna/US082B\_COMB.seq:\*

10: /cgn2\_6/ptodata/1/pna/US082C\_COMB.seq:\*

11: /cgn2\_6/ptodata/1/pna/US083A\_COMB.seq:\*

12: /cgn2\_6/ptodata/1/pna/US083B\_COMB.seq:\*

13: /cgn2\_6/ptodata/1/pna/US084A\_COMB.seq:\*

14: /cgn2\_6/ptodata/1/pna/US084B\_COMB.seq:\*

15: /cgn2\_6/ptodata/1/pna/US084C\_COMB.seq:\*

16: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq:\*

17: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq:\*

18: /cgn2\_6/ptodata/1/pna/US087A\_COMB.seq:\*

19: /cgn2\_6/ptodata/1/pna/US087B\_COMB.seq:\*

20: /cgn2\_6/ptodata/1/pna/US087C\_COMB.seq:\*

21: /cgn2\_6/ptodata/1/pna/US088A\_COMB.seq:\*

22: /cgn2\_6/ptodata/1/pna/US088B\_COMB.seq:\*

23: /cgn2\_6/ptodata/1/pna/US088C\_COMB.seq:\*

24: /cgn2\_6/ptodata/1/pna/US088D\_COMB.seq:\*

25: /cgn2\_6/ptodata/1/pna/US089A\_COMB.seq:\*

26: /cgn2\_6/ptodata/1/pna/US089B\_COMB.seq:\*

27: /cgn2\_6/ptodata/1/pna/US089C\_COMB.seq:\*

28: /cgn2\_6/ptodata/1/pna/US089D\_COMB.seq:\*

29: /cgn2\_6/ptodata/1/pna/US089E\_COMB.seq:\*

30: /cgn2\_6/ptodata/1/pna/US089F\_COMB.seq:\*

31: /cgn2\_6/ptodata/1/pna/US089G\_COMB.seq:\*

32: /cgn2\_6/ptodata/1/pna/US089H\_COMB.seq:\*

33: /cgn2\_6/ptodata/1/pna/US089I\_COMB.seq:\*

34: /cgn2\_6/ptodata/1/pna/US090A\_COMB.seq:\*

35: /cgn2\_6/ptodata/1/pna/US090B\_COMB.seq:\*

36: /cgn2\_6/ptodata/1/pna/US090C\_COMB.seq:\*

37: /cgn2\_6/ptodata/1/pna/US090D\_COMB.seq:\*

38: /cgn2\_6/ptodata/1/pna/US090E\_COMB.seq:\*

39: /cgn2\_6/ptodata/1/pna/US091A\_COMB.seq:\*

40: /cgn2\_6/ptodata/1/pna/US091B\_COMB.seq:\*

41: /cgn2\_6/ptodata/1/pna/US091C\_COMB.seq:\*

42: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq:\*

43: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq:\*

44: /cgn2\_6/ptodata/1/pna/US092C\_COMB.seq:\*

45: /cgn2\_6/ptodata/1/pna/US092D\_COMB.seq:\*

46: /cgn2\_6/ptodata/1/pna/US092E\_COMB.seq:\*

47: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq:\*

48: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq:\*

49: /cgn2\_6/ptodata/1/pna/US093C\_COMB.seq:\*

50: /cgn2\_6/ptodata/1/pna/US093D\_COMB.seq:\*

51: /cgn2\_6/ptodata/1/pna/US093E\_COMB.seq:\*

52: /cgn2\_6/ptodata/1/pna/US093F\_COMB.seq:\*

53: /cgn2\_6/ptodata/1/pna/US094A\_COMB.seq:\*

54: /cgn2\_6/ptodata/1/pna/US094B\_COMB.seq:\*

55: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq:\*

56: /cgn2\_6/ptodata/1/pna/US6001A\_COMB.seq:\*

57: /cgn2\_6/ptodata/1/pna/US6001B\_COMB.seq:\*

58: /cgn2\_6/ptodata/1/pna/US6002A\_COMB.seq:\*

59: /cgn2\_6/ptodata/1/pna/US6002B\_COMB.seq:\*

60: /cgn2\_6/ptodata/1/pna/US6002C\_COMB.seq:\*

61: /cgn2\_6/ptodata/1/pna/US6003A\_COMB.seq:\*

62: /cgn2\_6/ptodata/1/pna/US6003B\_COMB.seq:\*

63: /cgn2\_6/ptodata/1/pna/US6004A\_COMB.seq:\*

64: /cgn2\_6/ptodata/1/pna/US6004B\_COMB.seq:\*

65: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq:\*

66: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq:\*

67: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq:\*

68: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq:\*

69: /cgn2\_6/ptodata/1/pna/US6008A\_COMB.seq:\*

70: /cgn2\_6/ptodata/1/pna/US6008B\_COMB.seq:\*

71: /cgn2\_6/ptodata/1/pna/US6008C\_COMB.seq:\*

72: /cgn2\_6/ptodata/1/pna/US6009A\_COMB.seq:\*

73: /cgn2\_6/ptodata/1/pna/US6009B\_COMB.seq:\*

74: /cgn2\_6/ptodata/1/pna/US6010A\_COMB.seq:\*

75: /cgn2\_6/ptodata/1/pna/US6010B\_COMB.seq:\*

76: /cgn2\_6/ptodata/1/pna/US6011A\_COMB.seq:\*

77: /cgn2\_6/ptodata/1/pna/US6011B\_COMB.seq:\*

78: /cgn2\_6/ptodata/1/pna/US6012A\_COMB.seq:\*

79: /cgn2\_6/ptodata/1/pna/US6012B\_COMB.seq:\*

80: /cgn2\_6/ptodata/1/pna/US6013A\_COMB.seq:\*

81: /cgn2\_6/ptodata/1/pna/US6013B\_COMB.seq:\*

82: /cgn2\_6/ptodata/1/pna/US6014A\_COMB.seq:\*

83: /cgn2\_6/ptodata/1/pna/US6014B\_COMB.seq:\*

84: /cgn2\_6/ptodata/1/pna/US6014C\_COMB.seq:\*

85: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq:\*

86: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq:\*

87: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq:\*

88: /cgn2\_6/ptodata/2/pna/PCT\_NEW\_COMB.seq:\*

89: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB.seq:\*

90: /cgn2\_6/ptodata/2/pna/US07\_NEW\_COMB.seq:\*

91: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq:\*

92: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:\*

93: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result:	No.	Score	Query Match %	Length	DB	ID	Description
1	785	100.0	785	42	US-09-215-435-118		Sequence 118, Appl
2	785	100.0	785	69	US-60-081-363-71		Sequence 71, Appl
3	478	60.9	865	1	PCT-US99-13418-68		Sequence 68, Appl
4	393	50.1	928	34	US-09-009-325-13		Sequence 13, Appl
5	393	50.1	928	34	US-09-009-706-13		Sequence 13, Appl
6	393	50.1	928	34	US-09-009-802-13		Sequence 13, Appl
7	393	50.1	928	54	US-09-469-604-13		Sequence 13, Appl
8	393	50.1	928	92	US-09-009-685-13		Sequence 13, Appl
9	393	50.1	1232	92	US-09-491-404-3685		Sequence 3685, Ap
10	343	43.7	916	42	US-09-215-435-118		Sequence 311, Appl
11	343	43.7	916	67	US-60-069-957-144		Sequence 144, Appl
12	333	42.4	882	1	PCT-US99-13418-118		Sequence 118, Appl
13	325	41.4	464	25	US-08-905-279-38		Sequence 38, Appl

14	324	41.3	726	34	US-09-009-325-15	Sequence 15, Appl
15	324	41.3	726	34	US-09-009-706-15	Sequence 15, Appl
16	324	41.3	726	34	US-09-009-802-15	Sequence 15, Appl
17	324	41.3	726	34	US-09-469-604-15	Sequence 15, Appl
18	324	41.3	726	92	US-09-009-685-15	Sequence 15, Appl
19	235	29.9	354	45	US-09-289-768-25164	Sequence 25164, A
20	235	29.9	354	48	US-09-332-782-24407	Sequence 24407, A
21	101	12.9	273	39	US-09-105-427-2972	Sequence 2972, Ap
22	42	5.4	122	18	US-08-727-870-68	Sequence 68, Appl
23	42	5.4	122	55	US-60-005-023-68	Sequence 68, Appl
24	41	5.2	445	21	US-08-803-609-6177	Sequence 6177, Ap
25	39	5.0	333	13	US-08-401-881A-1563	Sequence 1563, Ap
26	39	5.0	333	13	US-08-401-881B-1563	Sequence 1563, Ap
27	34	4.3	364	13	US-08-401-881A-1590	Sequence 1590, Ap
28	34	4.3	364	13	US-08-401-881B-1590	Sequence 1590, Ap
c	29	33	425	92	US-09-515-126-13608	Sequence 13608, A
30	23	2.9	218	21	US-08-810-325-2993	Sequence 2993, Ap
31	23	2.9	218	56	US-60-012-458-470	Sequence 470, App
32	32	2.9	558	48	US-09-333-909-576	Sequence 576, App
c	33	22	486	93	US-60-189-657-14845	Sequence 14845, A
c	34	20	183	87	US-60-173-383-39816	Sequence 39816, A
35	20	2.5	678	93	US-60-182-467-387	Sequence 387, App
36	20	2.5	678	93	US-60-182-467-388	Sequence 388, App
37	20	2.5	678	93	US-60-182-467-389	Sequence 389, App
c	38	20	703	87	US-60-173-469-789	Sequence 789, App
c	39	20	1812	86	US-60-168-139-1288	Sequence 1288, Ap
c	40	20	2183	87	US-60-173-383-39815	Sequence 39815, A
c	41	20	10674	86	US-60-167-217-1577	Sequence 1577, Ap
c	42	20	10674	87	US-60-173-383-1645	Sequence 1645, Ap
c	43	20	10674	87	US-60-173-464-1291	Sequence 1291, Ap
c	44	20	22166	86	US-60-164-769-26164	Sequence 26164, A
c	45	20	22166	93	US-60-185-359-550	Sequence 550, App

RESULT 1

US-09-215-435-118

Sequence 118, Application US/09215435A

GENERAL INFORMATION:

APPLICANT: Dumas Milne Edwards, Jean-Baptiste

APPLICANT: Duclert Aymeric,

APPLICANT: Bougueleret Lydie

TITLE OF INVENTION: Extended cDNAs

FILE REFERENCE: GENSET.019A

CURRENT APPLICATION NUMBER: US/09/215.435A

CURRENT FILING DATE: 1998-12-17

EARLIER APPLICATION NUMBER: 60/069.957

EARLIER FILING DATE: 1997-12-17

EARLIER APPLICATION NUMBER: 60/074.121

EARLIER FILING DATE: 1998-2-9

EARLIER APPLICATION NUMBER: 60/081.563

EARLIER FILING DATE: 1998-4-13

EARLIER APPLICATION NUMBER: 60/096.116

EARLIER FILING DATE: 1998-8-10

NUMBER OF SEQ ID NOS: 519

SOFTWARE: Patent.pm

SEQ ID NO 118

LENGTH: 785

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: 72..704

FEATURE:

NAME/KEY: sig\_peptide

LOCATION: 72..161

OTHER INFORMATION: Von Heijne matrix

OTHER INFORMATION: score 13.2

OTHER INFORMATION: seq LLLSLTLVPSAA/AP

FEATURE:

NAME/KEY: polyA\_signal

US-09-215-435-118

LOCATION: 772..777

Query Match 100.0%; Score 785; DB 42; Length 785;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 785; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1

cggaaaccgggagtcggtgaccgggctggtgtagcataaaaggcgagccagaaga

60

DB 1

cggaaaccgggagtcggtgaccgggctggtgtagcataaaaggcgagccagaaga

60

QY 61

aggggggggtatggagaagcctccacactgcccgcgaagcgggatgtgtgtcc

120

DB 61

aggggggggtatggagaagcctccacactgcccgcgaagcgggatgtgtgtcc

120

QY 121

tgtctgtctctctctaccctggtgacccctcgctgcagctcctatccatgatgtg

180

DB 121

tgtctgtctctctctaccctggtgacccctcgctgcagctcctatccatgatgtg

180

QY 181

acgccaagagagctccttgggtctcacaggcctccagagctcactccaaggttcagcc

240

DB 181

acgccaagagagctccttgggtctcacaggcctccagagctcactccaaggttcagcc

240

QY 241

gactttctgaaagtgtaacctgttcggggcatagacagcttattctgtccccaagg

300

DB 241

gactttctgaaagtgtaacctgttcggggcatagacagcttattctgtccccaagg

300

QY 301

acttcggggcctcctgggaactaccacaaagagagaaaccagagcaccagctgggga

360

DB 301

acttcggggcctcctgggaactaccacaaagagagaaaccagagcaccagctgggga

360

QY 361

acaacacctctccagccacctccagatcgacaaggtaccaggtggagggaaggagg

420

DB 361

acaacacctctccagccacctccagatcgacaaggtaccaggtggagggaaggagg

420

QY 421

ccctgtaccatccagaagggcagggagagcttcacacagaaactccatccccgggtg

480

DB 421

ccctgtaccatccagaagggcagggagagcttcacacagaaactccatccccgggtg

480

QY 481

ccttctggatcattaaagctgccacggcgagggtcccacaggatgacctggaggcgcc

540

DB 481

ccttctggatcattaaagctgccacggcgagggtcccacaggatgacctggaggcgcc

540

QY 541

actggctcagagaagcagacccgctcagggcctcggagctcggagtgactcccgagggga

600

DB 541

actggctcagagaagcagacccgctcagggcctcggagctcggagtgactcccgagggga

600

QY 601

cccacaaggagctcctagaagaggggagcagagctcctccactccaggtgtccccc

660

DB 601

cccacaaggagctcctagaagaggggagcagagctcctccactccaggtgtccccc

660

QY 661

gaaagaccacttactgtatcatcctcagccctcctcggagctgtagggtgggaccgg

720

DB 661

gaaagaccacttactgtatcatcctcagccctcctcggagctgtagggtgggaccgg

720

QY 721

ggagcacctcctgtagccccctcagaccctgccccaaagcaccatattgaaataaagt

780

DB 721

ggagcacctcctgtagccccctcagaccctgccccaaagcaccatattgaaataaagt

780

QY 781

ctttc 785

|||||

DB 781

ctttc 785

|||||

RESULT 2

US-60-081-563-71

Sequence 71, Application US/60081563

GENERAL INFORMATION:

APPLICANT: Edwards, Jean-Baptiste Dumas Milne

APPLICANT: Duclert, Aymeric

TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SECRETED PROTEINS

NUMBER OF SEQUENCES: 129

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Win95
SOFTWARE: Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/60/081,563
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Israel, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: GENSET.027PR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 71:
SEQUENCE CHARACTERISTICS:
LENGTH: 785 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Testis
FEATURE:
NAME/KEY: CDS
LOCATION: 72..704
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 72..161
IDENTIFICATION METHOD: Von Heijne matrix
OTHER INFORMATION: score 13.2
OTHER INFORMATION: seq LLLSLTVIPSAA/AP
FEATURE:
NAME/KEY: poly_a_signal
LOCATION: 772..777
FEATURE:
NAME/KEY: est
LOCATION: 95..401
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 71..377
OTHER INFORMATION: id AA397836
FEATURE:
NAME/KEY: est
LOCATION: 61..103
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 38..80
OTHER INFORMATION: id AA397836
US-60-081-563-71

```

	Query Match	100.0%;	Score 785;	DB 69;	Length 785;	
	Best Local Similarity	100.0%;	Pred. No. 0;			
	Matches 785;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy	1	cggaaatcgggagtc	ccggtgacccgggctgtggttagcataaaggcggaagcccagaaga	.60		
Db	1	CGGAATCCGGAGTCCG	TGACCCGGGTGTGTTCTAGCATAAAGCGGAGGCCAAGA	60		
Qy	61	aggggcggggtatdggg	aagacctccccaccctgccccgcaaggcggaattgtcgtgtcc	120		
Db	61	AGGGGCGGGGTATGGG	AAGCCTCCCACCCTGCCCGCAAGCGGGCATCTGTGTGTCC	120		

Qy	121	tgtgctgcctctctctatacccttggtgatccccctccgctcgagctcctatcccatgatgctgy	180
Db	121	TGCTGTGCTGCCTCTCTACCCCTGGTGATCCCCCTCCGCTCAGCTCTATCCCATGATGCTG	180
Qy	181	acgcccaagaagactccttgggtctcacaggcctccagagcctactccaagctttcagcc	240
Db	181	ACGCCCAAGAAGACTCCTTGGGTCTCACAGCCTCCAGAGCCTTACTCCAAGGCTTCAGCG	240
Qy	241	gacttttcctgaaggttaaocctgtctcggggcagacagcgttatctctctgcccccattg	300
Db	241	GACTTTTCTTGAAGGTAACTCTTCCGGGCATACAGAGCTATTCTCTGCCCCCATTGG	300
Qy	301	acttcoggggcctcctcctggaaactaccacaagaagagagaaccaaggagcaccagctgggga	360
Db	301	ACTTCCGGGGCCTCCTCTGGGAATCCACAAGAGAGAGAACCCAGSAGACCAGCTGGGA	360
Qy	361	acaacacctctccagccacctccagatcgacaagtaccagagtaggagagagaagagg	420
Db	361	ACAACACCTCTCCAGCCACTCCAGATCGACAAGGTACCCAGGATGGAGGAGAAGAGG	420
Qy	421	cccttggtaccocatccagaagccacagcgacagcttccacacagacaactccccgggtgg	480
Db	421	CCCTGGTACCATCCAGAAGCCACAGACAGCTTCCACACAGAATCCATCCCCGGGTGG	480
Qy	481	cctctcgatcataagctcgcacgcggagaggtccacacagagatgcctctggaggcgccgc	540
Db	481	CCTTCTGGATCATTAAGCTGCCACGGCGGAGGTCCACACAGGATGCCCTGGAGGCGGCC	540
Qy	541	actgctccagcgagaagcgacacgcctcgcagcgcaatccgggagtgactccgcaagggga	600
Db	541	ACTGGCTCAGGAGAACGACACCGCCTCGAGGCCATCCGGGATGCACTCCGCAAGGGGA	600
Qy	601	cccacaaggagctoctagaaggggacogagagctctctccactccaggcttgtcccccc	660
Db	601	CCCAAGAGGAGCTCTAGAAGAGGAGGACCGAGAGAGCTCTCCCACTCCAGGCTGTCCCCC	660
Qy	661	gaagagcccactactgtacatcctcagggccctctcgagcagctgtgaggggtgggaccgg	720
Db	661	GAAAGACCCACTTACTGTACATCTCTAGGCCCTCTCGCAGCTGTAGGGGTGGGGACCG	720
Qy	721	ggagcacctgcctgtagccccccatcagaccctgccccaaagcaccatatggaataaagtt	780
Db	721	GGAGCACCTGCTGTAGCCCCCATCAGACCCCTGCCCAAGCACCATATGAAATAAAGTT	780
Qy	781	ctttc 785	
Db	781	ctttc 785	
RESULT 3			
PCT-US99-13418-68			
; Sequence 68, Application PC/TUS9913418			
; GENERAL INFORMATION:			
; APPLICANT: Human Genome Sciences, Inc. et al.			
; TITLE OF INVENTION: 94 Human secreted proteins			
; FILE REFERENCE: P2029PCT			
; CURRENT APPLICATION NUMBER: PCT/US99/13418			
; CURRENT FILING DATE: 1999-06-15			
; EARLIER APPLICATION NUMBER: 60/089,508			
; EARLIER FILING DATE: 1998-06-16			
; EARLIER APPLICATION NUMBER: 60/089,507			
; EARLIER FILING DATE: 1998-06-16			
; EARLIER APPLICATION NUMBER: 60/089,510			
; EARLIER FILING DATE: 1998-06-16			
; EARLIER APPLICATION NUMBER: 60/089,509			
; EARLIER FILING DATE: 1998-06-16			
; EARLIER APPLICATION NUMBER: 60/090,112			
; EARLIER FILING DATE: 1998-06-22			
; EARLIER APPLICATION NUMBER: 60/090,113			
; EARLIER FILING DATE: 1998-06-22			
; NUMBER OF SEQ ID NOS: 502			
; SOFTWARE: PatentIn Ver. 2.0			

; SEQ ID NO 68  
; LENGTH: 865  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (445)  
; OTHER INFORMATION: n equals a,t,g, or c  
PCT-US99-13418-68

Query Match 60.9%; Score 478; DB 1; Length 865;  
Best Local Similarity 99.3%; Pred. No. 1.7e-225;  
Matches 728; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 53 ccagaaagggcggtatggagagctcccccacccgcccccaaggcggtatct 112  
Db 118 ccagaaagggcggtatggagagctcccccacccgcccccaaggcggtatct 177  
QY 113 gctggtctgctgctctctctacccgtggtgacccctccgctgcagctccatcca 172  
Db 178 gctggtctgctgctctctctacccgtggtgacccctccgctgcagctccatcca 237  
QY 173 tgatgctagcccaagagactcctgggtctcacaggccctcagagcctactccaagg 232  
Db 238 tgatgctagcccaagagactcctgggtctcacaggccctcagagcctactccaagg 297  
QY 233 cttagccgacttttctgaaaggttaacctgtctcggggcatagacagcttattctctgc 292  
Db 298 cttagccgacttttctgaaaggttaacctgtctcggggcatagacagcttattctctgc 357  
QY 293 ccccatgacttcggggtcctctctggaactaccacaaagagagaccaggagaccacca 352  
Db 358 ccccatgacttcggggtcctctctggaactaccacaaagagagaccaggagaccacca 417  
QY 353 gctggggaacacacctctcagccacctcagatgacagatgacagatggagga 412  
Db 418 rctggggaacacacctctcagccacctcagatgacagatgacagatggagga 477  
QY 413 gaaggagccctgttaccctccagaaagccacgacagcttccacacagactccatcc 472  
Db 478 gaaggagccctgttaccctccagaaagccacgacagcttccacacagactccatcc 537  
QY 473 ccgggtgctcttgatcataagctccacgcgagggtccacacagagatgacctgga 532  
Db 538 ccgggtgctcttgatcataagctccacgcgagggtccacacagagatgacctgga 597  
QY 533 gggcgccactggctcagcgaagcgcacacgctcagggccatccggggtgactccg 592  
Db 598 gggcgccactggctcagcgaagcgcacacgctcagggccatccggggtgactccg 657  
QY 593 caaggggaccacaaagagcgtctctagaaaggggacagagagctctccactccaggct 652  
Db 658 caaggggaccacaaagagcgtctctagaaaggggacagagagctctccactccaggct 717  
QY 653 gtccccccaaagacccactactatcatctcagagccctcgcagctgtagggtg 712  
Db 718 gtccccccaaagacccactactgtactatctcagagccctcgcagctgtagggtg 777  
QY 713 gggagccgggagcactgctgtagcccccatcagaccctgccccaaagcaccatggaa 772  
Db 778 gggagccgggagcactgctgtagcccccatcagaccctgccccaaagcaccatggaa 837  
QY 773 ataaagtcttttc 785  
Db 838 ataaagtcttttc 850

RESULT 4  
US-09-009-325-13  
; Sequence 13, Application US/09009325A  
; GENERAL INFORMATION:  
; APPLICANT: McCarthy, Sean A.

; TITLE OF INVENTION: NOVEL CRSP PROTEIN AND NUCLEIC ACID MOLECULES AND USES  
; TITLE OF INVENTION: THEREFOR  
; FILE REFERENCE: MEI-008DV3  
; CURRENT APPLICATION NUMBER: US/09/009,325A  
; CURRENT FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/842,898  
; EARLIER FILING DATE: 1997-04-17  
; EARLIER APPLICATION NUMBER: 60/071,589  
; EARLIER FILING DATE: 1998-01-15  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 928  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (75)...(800)  
US-09-009-325-13

Query Match 50.1%; Score 393; DB 34; Length 928;  
Best Local Similarity 100.0%; Pred. No. 1.6e-183;  
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 aaggtaccagagtgagagagagggccctggtaccatccagaaagccacggacagc 452  
Db 489 aaggtaccagagtgagagagagggccctggtaccatccagaaagccacggacagc 548  
QY 453 ttccacacagaactccatccccgggtggtctctggtatcataagctgccacggcgag 512  
Db 549 ttccacacagaactccatccccgggtggtctctggtatcataagctgccacggcgag 608  
QY 513 tcccaccagagtcctctgagggcgccactggtcagcgagagcgacacccctcgag 572  
Db 609 tcccaccagagtcctctgagggcgccactggtcagcgagagcgacacccctcgag 668  
QY 573 gcataccggatgagtcctcgaagggggacccacaaagagctctagaagaggggacccag 632  
Db 669 gccatccggatgagtcctcgaagggggacccacaaagagctctagaagaggggacccag 728  
QY 633 agctctccacatccagctgtctcccccgaagagacccacttactgtacatctcagggccc 692  
Db 729 agctctccacatccagctgtctcccccgaagagacccacttactgtacatctcagggccc 788  
QY 693 tctcggcagctgtagggtggtgggacccgggagacactgctgtagcccccatcagaccct 752  
Db 789 tctcggcagctgtagggtggtgggacccgggagacactgctgtagcccccatcagaccct 848  
QY 753 gccccagcaccataggaaataaagtctttc 785  
Db 849 gccccagcaccataggaaataaagtctttc 881

RESULT 5  
US-09-009-706-13  
; Sequence 13, Application US/09009706A  
; GENERAL INFORMATION:  
; APPLICANT: McCarthy, Sean A.  
; TITLE OF INVENTION: NOVEL CRSP PROTEIN AND NUCLEIC ACID MOLECULES AND USES  
; TITLE OF INVENTION: THEREFOR  
; FILE REFERENCE: MEI-008DV1  
; CURRENT APPLICATION NUMBER: US/09/009,706A  
; CURRENT FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/842,898  
; EARLIER FILING DATE: 1997-04-17  
; EARLIER APPLICATION NUMBER: 60/071,589  
; EARLIER FILING DATE: 1998-01-15  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 928  
; TYPE: DNA



```
|||||
Db 669 gcatccggatggaccccgcaaggaggacacgacgtcctagaaaggggacccag 728
QY 633 agctctccactccagctgtcccccgaagaccacttactgtacatctccagccc 692
|||||
Db 729 agctctccactccagctgtcccccgaagaccacttactgtacatctccagccc 788
QY 693 tctcggagctgtagggtgggagaccgggagacactgctgtagccccatcagacct 752
|||||
Db 789 tctcggagctgtagggtgggagaccgggagacactgctgtagccccatcagacct 848
QY 753 gccccagaccatggaataaagtctttc 785
|||||
Db 849 gccccagaccatggaataaagtctttc 881

RESULT 8
US-09-009-685-13
; Sequence 13, Application US/09009685A
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: NOVEL CRSP PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: MEI-008DV2
; CURRENT APPLICATION NUMBER: US/09/009,685A
; CURRENT FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/843,704
; EARLIER FILING DATE: 1997-04-16
; EARLIER APPLICATION NUMBER: 08/842,898
; EARLIER FILING DATE: 1997-04-17
; EARLIER APPLICATION NUMBER: 60/071,589
; EARLIER FILING DATE: 1998-01-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 928
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (75)..(800)
US-09-009-685-13

Query Match 50.1%; Score 393; DB 92; Length 928;
Best Local Similarity 100.0%; Pred. No. 1.6e-183;
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 aaggtaccagatggaggaaggagccctgtaccatccagagggccagcagc 452
|||||
Db 489 aaggtaccagatggaggaaggagccctgtaccatccagagggccagcagc 548
QY 453 ttccacacagaactccatcccccgggtgcttctggtacattaaagctgccacggcgagg 512
Db 549 ttccacacagaactccatcccccgggtgcttctggtacattaaagctgccacggcgagg 608
QY 513 ttccacacagaactccatcccccgggtgcttctggtacattaaagctgccacggcgagg 572
Db 609 ttccacacagaactccatcccccgggtgcttctggtacattaaagctgccacggcgagg 668
QY 573 gcatccggagatgactccgcaaggaccacacagacactcctagaaggggacccag 632
Db 669 gcatccggagatgactccgcaaggaccacacagacactcctagaaggggacccag 728
QY 633 agctctccactccagctgtcccccgaagaccacttactgtacatctccagccc 692
Db 729 agctctccactccagctgtcccccgaagaccacttactgtacatctccagccc 788
QY 693 tctcggagctgtagggtgggagaccgggagacactgctgtagccccatcagacct 752
Db 789 tctcggagctgtagggtgggagaccgggagacactgctgtagccccatcagacct 848
QY 753 gccccagaccatggaataaagtctttc 785
```

```
|||||
Db 849 gccccagaccatggaataaagtctttc 881

RESULT 9
US-09-491-404-3685
; Sequence 3685, Application US/09491404
; GENERAL INFORMATION:
; APPLICANT: Tang, Yuanhua T.
; APPLICANT: Tillinghast, John
; APPLICANT: Sinku, Ankura
; APPLICANT: Liu, Chenghua
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Contigs Obtained
; TITLE OF INVENTION: From Various Libraries
; FILE REFERENCE: 785
; CURRENT APPLICATION NUMBER: US/09/491,404
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 3796
; SOFTWARE: pt_SP_genes Version 1.0
; SEQ ID NO 3685
; LENGTH: 1232
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (406)...(541)
; OTHER INFORMATION: This location contains the signal peptide sequence,
; OTHER INFORMATION: MGEASPPAPARRHLLVLLLLSTIVTSPAAPIHDADAQSSLSL, Run with signal
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (406)...(801)
; OTHER INFORMATION: Similar to g1517093 in the genepept database release 114,
; OTHER INFORMATION: Run with FASTXY 3.3t00, default parameters
US-09-491-404-3685

Query Match 50.1%; Score 393; DB 92; Length 1232;
Best Local Similarity 100.0%; Pred. No. 1.6e-183;
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 aaggtaccagatggaggaaggagccctgtaccatccagagggccagcagc 452
|||||
Db 820 aaggtaccagatggaggaaggagccctgtaccatccagagggccagcagc 879
QY 453 ttccacacagaactccatcccccgggtgcttctggtacattaaagctgccacggcgagg 512
Db 880 ttccacacagaactccatcccccgggtgcttctggtacattaaagctgccacggcgagg 939
QY 513 ttccacacagaactccatcccccgggtgcttctggtacattaaagctgccacggcgagg 572
Db 940 ttccacacagaactccatcccccgggtgcttctggtacattaaagctgccacggcgagg 999
QY 573 gcatccggagatgactccgcaaggaccacacagacactcctagaaggggacccag 632
Db 1000 gcatccggagatgactccgcaaggaccacacagacactcctagaaggggacccag 1059
QY 633 agctctccactccagctgtcccccgaagaccacttactgtacatctccagccc 692
Db 1060 agctctccactccagctgtcccccgaagaccacttactgtacatctccagccc 1119
QY 693 tctcggagctgtagggtgggagaccgggagacactgctgtagccccatcagacct 752
Db 1120 tctcggagctgtagggtgggagaccgggagacactgctgtagccccatcagacct 1179
QY 753 gccccagaccatggaataaagtctttc 785
Db 1180 gccccagaccatggaataaagtctttc 1212

RESULT 10
US-09-215-435-311
; Sequence 311, Application US/09215435A
```



```

; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric,
; APPLICANT: Bouquelieret Lydie
; TITLE OF INVENTION: Extended cDNAs
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215,435A
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/069,957
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 311
; LENGTH: 916
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 90..815
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 90..179
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 13.199998092651
; OTHER INFORMATION: seq LLLUSTLVIPSAAP/AP
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 883..888
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 905..916
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 487,661
; OTHER INFORMATION: n=a, g, c or t
US-09-215-435-311

```

```

Query Match      43.7%; Score 343; DB 42; Length 916;
Best Local Similarity 100.0%; Pred. No. 8e-159;
Matches 343; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 ccagaagaagggcggtatggagaagcctccacacgtgcccgcaaggcgcatct 112
Db 71 ccagaagaagggcggtatggagaagcctccacacgtgcccgcaaggcgcatct 130
QY 113 gctggtcctgctgctcctctctacccctggtgtagccctccgctgcagctcctatcca 172
Db 131 gctggtcctgctgctcctctctacccctggtgtagccctccgctgcagctcctatcca 190
QY 173 tgatgctgacgcccaagaagatccttgggtctcagcgcctccagagcctactccaagg 232
Db 191 tgatgctgacgcccaagaagatccttgggtctcagcgcctccagagcctactccaagg 250
QY 233 ctccagccgactttcctgaaaggttaacctgttcgggagctagacagcttattctctgc 292
Db 251 ctccagccgactttcctgaaaggttaacctgttcgggagctagacagcttattctctgc 310
QY 293 ccccatgggacttcggggcctccctgggaactaccacaagaggagaaaccaggagaccaca 352
Db 311 ccccatgggacttcggggcctccctgggaactaccacaagaggagaaaccaggagaccaca 370
QY 353 gctggggaacaacacctctccagccacctccagatcgacaag 395
Db 371 gctggggaacaacacctctccagccacctccagatcgacaag 413

```

```

RESULT 11
US-60-069-957-144
; Sequence 144, Application US/60069957
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste Dumas Milne
; APPLICANT: Duclert, Aymeric
; TITLE OF INVENTION: EXTENDED cDNAs FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 381
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/069,957
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.019PR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 144:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 916 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Testis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 90..815
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 90..179
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 13.2
; OTHER INFORMATION: seq LLLUSTLVIPSAAP/AP
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 883..888
; FEATURE:
; NAME/KEY: poly_a
; LOCATION: 905..916
; IDENTIFICATION METHOD: blastn
; FEATURE:
; NAME/KEY: est
; LOCATION: 113..482
; IDENTIFICATION METHOD: blastn
; OTHER INFORMATION: identity 97
; OTHER INFORMATION: region 71..440
; OTHER INFORMATION: id AA397836
; FEATURE:
; NAME/KEY: est
; LOCATION: 79..121
; IDENTIFICATION METHOD: blastn
; OTHER INFORMATION: identity 100
; OTHER INFORMATION: region 38..80
; OTHER INFORMATION: id AA397836
US-60-069-957-144

```

Query Match 43.7%; Score 343; DB 67; Length 916;  
 Best Local Similarity 100.0%; Pred. No. 8e-159;  
 Matches 343; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 ccagaagaaggggggtatggagaaagcctcccccacatgccccgcgcaaggcggtatct 112  
 |||  
 Db 71 CCAGAAAGAGGGGGGTATGGAGAGCCTCCACACCTGCCCCGCGCAAGGGCGCATCT 130  
 |||

QY 113 gctggtcctca 172  
 |||  
 Db 131 GCTGGGCTCA 190  
 |||

QY 173 tgatgctgagcccaagagagctcccttgggtctcacaggcctccagagcctactccaagg 232  
 |||  
 Db 191 TGATGCTGAGCCCAAGAGAGCTCTTGGTCTCTACAGGCTCTACAGGCTACTCCAAG 250  
 |||

QY 233 cttcagccgacttttctgaaaggttaacctgtctcgggcatagacagcttattctctgc 292  
 |||  
 Db 251 CTTACGCCGACTTTTCTGAAAGGTAACTCTCTGGGCTCTGCGGCTATAGACAGCTTATCTCTGC 310  
 |||

QY 293 cccatggacttccgggctcctcctgggaactaccacaagagagagacagagcacca 352  
 |||  
 Db 311 CCCATGGACTTCCGGGGCTCTCTGGAACTACTACCAAGAGAGAGACAGGAGCACCA 370  
 |||

QY 353 gctggggaacaacacccctctccagccacctccagatcgacaag 395  
 |||  
 Db 371 GCTGGGAACAACACCTCTCCAGCCACCTCCAGATCGACAAG 413  
 |||

RESULT 12  
 PCT-US99-13418-118  
 ; Sequence 118, Application PC/TUS9913418  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Human Genome Sciences, Inc. et al.  
 ; TITLE OF INVENTION: 94 Human secreted proteins  
 ; FILE REFERENCE: P2029PCT  
 ; CURRENT APPLICATION NUMBER: PCT/US99/13418  
 ; CURRENT FILING DATE: 1999-06-15  
 ; EARLIER APPLICATION NUMBER: 60/089,508  
 ; EARLIER FILING DATE: 1998-06-16  
 ; EARLIER APPLICATION NUMBER: 60/089,507  
 ; EARLIER FILING DATE: 1998-06-16  
 ; EARLIER APPLICATION NUMBER: 60/089,510  
 ; EARLIER FILING DATE: 1998-06-16  
 ; EARLIER APPLICATION NUMBER: 60/089,509  
 ; EARLIER FILING DATE: 1998-06-16  
 ; EARLIER APPLICATION NUMBER: 60/090,112  
 ; EARLIER FILING DATE: 1998-06-22  
 ; EARLIER APPLICATION NUMBER: 60/090,113  
 ; EARLIER FILING DATE: 1998-06-22  
 ; NUMBER OF SEQ ID NOS: 502  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 118  
 ; LENGTH: 882  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (117)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 PCT-US99-13418-118

Query Match 42.4%; Score 333; DB 1; Length 882;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-154;  
 Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 aaggtaccagagtgaggagc 452  
 |||  
 Db 550 aaggtaccagagtgaggagagagagagagagagagagagagagagagagagagagc 609  
 |||

QY 453 ttccacagaaactccatcccccgggtgctcttgatcattagctgccacggcgagg 512

Db 610 ttccacagaaactccatcccccgggtgctcttgatcattagctgccacggcgagg 669  
 |||

QY 513 tccacacagagtgagggcgagggcgagggcgagggcgagggcgagggcgagggcgag 572  
 |||

Db 670 tccacacagagtgagggcgagggcgagggcgagggcgagggcgagggcgagggcgag 729  
 |||

QY 573 gcatcccgagtgagggcgagggcgagggcgagggcgagggcgagggcgagggcgag 632  
 |||

Db 730 gcatcccgagtgagggcgagggcgagggcgagggcgagggcgagggcgagggcgag 789  
 |||

QY 633 agctcctccacactcagagctgtcccccgagggcgagggcgagggcgagggcgagggcgag 692  
 |||

Db 790 agctcctccacactcagagctgtcccccgagggcgagggcgagggcgagggcgagggcgag 849  
 |||

QY 693 tctcgagagctgtaggggtgaggggagggcgagggcgagggcgagggcgagggcgag 725  
 |||

Db 850 tctcgagagctgtaggggtgaggggagggcgagggcgagggcgagggcgagggcgag 882  
 |||

RESULT 13  
 US-08-905-279-38  
 ; Sequence 38, Application us/08905279  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Genset SA  
 ; TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS  
 ; TITLE OF INVENTION: EXPRESSED IN TESTIS AND OTHER TISSUES  
 ; NUMBER OF SEQUENCES: 503  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear  
 ; STREET: 501 West Broadway  
 ; CITY: San Diego  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 92101-3505  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy Disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: Win95  
 ; SOFTWARE: Word  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/905,279  
 ; FILING DATE:  
 ; CLASSIFICATION: 536  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Israelsen, Ned A.  
 ; REGISTRATION NUMBER: 29,655  
 ; REFERENCE/DOCKET NUMBER: GENSET.011A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (619) 235-8550  
 ; TELEFAX: (619) 235-0176  
 ; INFORMATION FOR SEQ ID NO: 38:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 464 base pairs  
 ; TYPE: NUCLEIC ACID  
 ; STRANDEDNESS: DOUBLE  
 ; TOPOLOGY: LINEAR  
 ; MOLECULE TYPE: CDNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Homo Sapiens  
 ; TISSUE TYPE: Testis  
 ; FEATURE:  
 ; NAME/KEY: sig\_peptide  
 ; LOCATION: 90..179  
 ; IDENTIFICATION METHOD: Von Heijne matrix  
 ; OTHER INFORMATION: score 13.2  
 ; OTHER INFORMATION: seq LILLSTLVPISAA/AP  
 US-08-905-279-38

Query Match 41.4%; Score 325; DB 25; Length 464;  
 Best Local Similarity 100.0%; Pred. No. 6.2e-150;  
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 53 ccagaagaaggggggtatgggagaagcctccccacactgcccccgcaagcgccatct 112
Db 71 CCAGAAGAAGGGGGGTATGGGAGAAGCTCCACACTGCCCGCGCAAGCGGCATCT 130
Qy 113 gctggtctgtgtgtctctctctacactgtgtgacccctccgctcagctctctatcca 172
Db 131 GCTGTCTGTGTGTGTCTCTCTCTACCTGTGTATCCCTCCGCTGAGTCTCTATCCA 190
Qy 173 tgatgtacgcccagaagagctcttgggtctcacaggctccagagcctactcccaagg 232
Db 191 TGATGCTGACGCCCAAGAGAGCTCTTGGGTCTCACAGGCTCCAGAGCTTCCRAAG 250
Qy 233 ctccagcgaatttctctgaagtaacctgtcttggggatagacagcttattctctgc 292
Db 251 CTTGAGCGGACTTTTCTCTGAAGGTAACCTGCTTGGGGCATAGACAGCTTATTCTCTGC 310
Qy 293 ccccatgacttcggggcctccctcctgggaactaccacaaagagagaccagagaccaca 352
Db 311 CCCATGAGACTTCGGGGCCCTCCCTGGGAATCTACCAAGAGGAGAACAGGAGCACCA 370
Qy 353 gctgggggaacacacctctccagc 377
Db 371 GCTGGGGAACACACCTCTCCAGC 395

```

```

RESULT 14
US-09-009-325-15
; Sequence 15, Application US/09009325A
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: NOVEL CRSP PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: MEI-008DV3
; CURRENT APPLICATION NUMBER: US/09/009,325A
; CURRENT FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/842,898
; EARLIER FILING DATE: 1997-04-17
; EARLIER APPLICATION NUMBER: 60/071,589
; EARLIER FILING DATE: 1998-01-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-09-009-325-15

```

```

Query Match 41.3%; Score 324; DB 34; Length 726;
Best Local Similarity 100.0%; Pred. No. 1.9e-149;
Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 72 atgggagaagcctccccacactgcccccgcaagcgccatctgtgtgtctgtctgtc 131
Db 1 atgggagaagcctccccacactgcccccgcaagcgccatctgtgtgtctgtctgtc 60
Qy 132 ctctctacctgtgtgacccctccgctcagctctctatccatgatgctgagccccaaag 191
Db 61 ctctctacctgtgtgacccctccgctcagctctctatccatgatgctgagccccaaag 120
Qy 192 agctccttgggtctcacaggctccagagcctactccaaggcttcagccgactttctctg 251
Db 121 agctccttgggtctcacaggctccagagcctactccaaggcttcagccgactttctctg 180
Qy 252 aaaggaacactgcttcggggcatagacagcttattctctgccccatggacttcggggc 311
Db 181 aaaggaacactgcttcggggcatagacagcttattctctgccccatggacttcggggc 240
Qy 312 ctccctgggaactaccacaaagagagaccagagcaccagctggtgggaacacacccctc 371
Db 241 ctccctgggaactaccacaaagagagaccagagcaccagctggtgggaacacacccctc 300
Qy 372 tccagccacctccagatcgacaag 395
Db 301 tccagccacctccagatcgacaag 324

```

Search completed: May 1, 2000, 20:00:14  
Job time: 21998 sec

```

Db 241 ctccctgggaactaccacaaagagagaccagagcaccagctggtgggaacacacccctc 300
Qy 372 tccagccacctccagatcgacaag 395
Db 301 tccagccacctccagatcgacaag 324

RESULT 15
US-09-009-706-15
; Sequence 15, Application US/09009706A
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: NOVEL CRSP PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: MEI-008DV1
; CURRENT APPLICATION NUMBER: US/09/009,706A
; CURRENT FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/842,898
; EARLIER FILING DATE: 1997-04-17
; EARLIER APPLICATION NUMBER: 60/071,589
; EARLIER FILING DATE: 1998-01-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-09-009-706-15

```

```

Query Match 41.3%; Score 324; DB 34; Length 726;
Best Local Similarity 100.0%; Pred. No. 1.9e-149;
Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 72 atgggagaagcctccccacactgcccccgcaagcgccatctgtgtgtctgtctgtc 131
Db 1 atgggagaagcctccccacactgcccccgcaagcgccatctgtgtgtctgtctgtc 60
Qy 132 ctctctacctgtgtgacccctccgctcagctctctatccatgatgctgagccccaaag 191
Db 61 ctctctacctgtgtgacccctccgctcagctctctatccatgatgctgagccccaaag 120
Qy 192 agctccttgggtctcacaggctccagagcctactccaaggcttcagccgactttctctg 251
Db 121 agctccttgggtctcacaggctccagagcctactccaaggcttcagccgactttctctg 180
Qy 252 aaaggaacactgcttcggggcatagacagcttattctctgccccatggacttcggggc 311
Db 181 aaaggaacactgcttcggggcatagacagcttattctctgccccatggacttcggggc 240
Qy 312 ctccctgggaactaccacaaagagagaccagagcaccagctggtgggaacacacccctc 371
Db 241 ctccctgggaactaccacaaagagagaccagagcaccagctggtgggaacacacccctc 300
Qy 372 tccagccacctccagatcgacaag 395
Db 301 tccagccacctccagatcgacaag 324

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 14:44:45 ; Search time 3022.95 Seconds  
(without alignments)  
980.465 Million cell updates/sec

Title: US-09-215-435-118  
Perfect score: 785  
Sequence: 1 cggaaatccggagtcggtg.....tatggaataaaagtcttttc 785

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4538634 seqs, 1887831982 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*

44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_gss1:\*  
80: gb\_gss2:\*  
81: gb\_gss3:\*  
82: gb\_gss4:\*  
83: em\_gss1:\*  
84: em\_gss2:\*  
85: em\_gss3:\*  
86: em\_gss4:\*  
87: gb\_gss5:\*  
88: gb\_gss6:\*  
89: gb\_gss7:\*  
90: gb\_gss8:\*  
91: gb\_gss9:\*  
92: em\_gss5:\*  
93: em\_gss6:\*  
94: em\_gss7:\*  
95: em\_gss8:\*  
96: em\_gss9:\*  
97: em\_gss10:\*  
98: em\_gss11:\*  
99: gb\_gss10:\*  
100: gb\_gss11:\*  
101: em\_gss12:\*  
102: gb\_gss12:\*  
103: gb\_gss13:\*  
104: gb\_gss14:\*  
105: gb\_gss15:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	393	50.1	432	43	AI200822
					AI200822 qf67e03.x



ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 483)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 14, 1998 this sequence version replaced gi:1796891.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
 Insert Length: 982 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amer-sham  
 High quality sequence stop: 414.  
 Location/Qualifiers  
 1..483  
 /organism="Homo sapiens"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from Clontech  
 Laboratories, Inc., and primed with a Not I - oligo(dT)  
 primer [5'  
 TGTACCACTCTGAAGTGGGAGCGCGCCCAATTTTTTTTTTTTT 3']  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization to Cot5, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."  
 77 a 143 c 148 g 115 t  
 BASE COUNT  
 ORIGIN  
 Query Match 50.1%; Score 393; DB 41; Length 483;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-190;  
 Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 393 aaggtaccagatgagagagagagccctgtaccatccagaagccacgacagc 452  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||  
 406 AAGGTACCCAGGATGAGGAGAGAGGCGCCCTGGTACCATCCAGAAGCCACGGACACG 347  
 QY |||||||||||||||||||||||||||||||||||||||||||||||||||  
 453 ttcacacagaactccatcccggtggtcctctggatcattaaagtcgacgagcgag 512  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||  
 346 TTCCACACAGAAGTCCATCCCGGGTGGCTTCTGGATCATTAAGCTGCACGGCGAGG 287  
 QY |||||||||||||||||||||||||||||||||||||||||||||||||||  
 513 tccaccagatgctgagagcgccactgctcagcagagcagacccgctcgag 572  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||  
 286 TCCACACAGGATGCTTGAGAGGGCGGCCACTGGCTCAGCGAAGACGACCGCTGCAG 227  
 QY |||||||||||||||||||||||||||||||||||||||||||||||||||  
 573 gccatccgggatggactccgcgaaggggacccacaaggagcgtcctagaagagggaccag 632  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||  
 226 GCATCCGGGATGGACTCCGCAAGGGGACCCCAAGGAGCTCTAGAAGAGGGGACCGAG 167  
 QY |||||||||||||||||||||||||||||||||||||||||||||||||||  
 633 agctctcccatccaggtgttccccccgaagacccacttactgtacatctcagggccc 692  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||  
 166 AGCTCCTCCACATCCAGGCTGTCCCCCGAAGAGCCCACTTACTGTACATCTCAGGCC 107  
 QY |||||||||||||||||||||||||||||||||||||||||||||||||||  
 693 tctcgagcgtatgggggtgggacgggagcactgctgtagcccccatcagacct 752  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||  
 106 TCTGCGACGTGTAGGGGTGGGAGCGGGGAGCACCTGCTGTAGCCCCCATCAGACCT 47

QY 753 gccccaagcaccatattgaaataaagtctttc 785  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||  
 46 GCCCAAGCACCATATGGAATAAAGTCTTCTTC 14  
 RESULT 3  
 AW162347/c 598 bp mRNA EST 09-NOV-1999  
 LOCUS au74c11.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone  
 DEFINITION IMAGE:2782004 3', similar to SW:LFE4.CHICK Q90839 UNKNOWN LENS FIBER  
 PROTEIN CLFE4 PRECURSOR. ; mRNA sequence.  
 ACCSSION AW162347 GI:5301380  
 VERSION AW162347.1  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 598)  
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
 Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
 WashU-NCI human EST Project  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 5, 1998 this sequence version replaced gi:3189341.  
 Other ESTs: au74c11.y1  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 464.  
 Location/Qualifiers  
 1..598  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2782004"  
 /clone\_lib="Schneider fetal brain 00004"  
 /sex="male"  
 /tissue\_type="frontal lobe"  
 /dev\_stage="5 months post-conception"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pBluescript SK (Stratagene);  
 Site\_1: SstI; Site\_2: XhoI; Double-stranded cDNA was  
 prepared from human fetal brain tissue. 5' and 3'  
 adaptors were used in cloning as follows: 5' adaptor  
 sequence:  
 5'-GAGAGAGAGAGACTCAAGGATCCTTAATTAATTAATTCCTCCCCCCCC-3'  
 and 3' adaptor sequence:  
 5'-GAGAGAGAGACTCGAGTGTGTGTGTGTGTGTGTGT-3'. The library was  
 size-selected for >0.5 kb inserts and has an average  
 insert size estimated at 1.2 kb. This library was  
 constructed using the CAP-trapper method for full-length  
 enrichment and has not undergone amplification. Library  
 was constructed by Dr. Claudio Schneider (LNCIB-Area  
 Science Park, Trieste, Italy)."  
 91 a 168 c 184 g 155 t  
 BASE COUNT  
 ORIGIN  
 Query Match 50.1%; Score 393; DB 72; Length 598;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-190;  
 Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 393 aaggtaccagatgagagagagccctgtaccatccagaagccacgacagc 452  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||  
 409 AAGGTACCCAGGATGAGGAGAGGCGCCCTGGTACCATCCAGAAGCCACGGACAGC 350

QY 453 ttccacagaactcattcccccgggtggtcttcttgatattaaagctgcacggcgagg 512  
 Db 349 TTCCACACAGAACTCCATCCCGGGTGGGCTTCTGGATCATTAAGCTGCCACGGCGAGG 290  
 QY 513 tccacacagatcccttgaggcgccactggtcagcgagaagcgacacgcctgcag 572  
 Db 289 TCCACACAGGATGCTTGGAGGGCGCCACTGGCTCAGCGAGAAGCGACCGCTGCAG 230  
 QY 573 gcatccggatgagctccgcgaaggagacccacaaagagcgtctagaagagggaccgag 632  
 Db 229 GCATCCGGGATGAGTCCGCAAGGGAGCCCAAGAGAGCTCTAGAGAGGGAGCCGAG 170  
 QY 633 agctctccactccaggtgtccccccaaagaccacttactgtacatctcagggccc 692  
 Db 169 AGCTCTCCACCTCCAGCTGTCCCCCGGAGACCCACTTACTGTACATCTCAGGCC 110  
 QY 693 tctcggcagctgtaggggtgggacggggagcacctgctgtagcccccatcagacct 752  
 Db 109 TCTCGGAGCTGTAGGGTGGGACGGGAGCACCTGCTGTAGCCCCCATCAGACCT 50  
 QY 753 gccccaaagcattatggaataaagtctttc 785  
 Db 49 GCCCCAAGCACCATTATGGAATAAAGTCTTTC 17

RESULT 4  
 AI150592/c  
 LOCUS  
 DEFINITION  
 3', mRNA sequence.  
 ACCESSION  
 AI150592  
 VERSION  
 AI150592.1  
 KEYWORDS  
 EST.  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 428)  
 AUTHORS  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL  
 Unpublished (1997)  
 COMMENT  
 On Sep 12, 1996 this sequence version replaced gi:1397999.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone Distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 506 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 419.  
 Location/Qualifiers  
 1. 428  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:1752164"  
 /clone\_lib="Soares\_testis\_NHT"  
 /sex="male"  
 /lab\_host="DH10B"  
 /note="Vector: p7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from Clontech  
 Laboratories, Inc., and primed with a Not I - oligo(dT)  
 primer [5].  
 TGTATCAATCTGAAGTGGGAGCGCGCCCAATTTTTTTTTTTTTT 3'.  
 Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified p7T3 vector. Library  
 went through one round of normalization to Cot5, and was  
 constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 67 a 126 c 135 g 99 t 1 others

Query Match 49.0%; Score 385; DB 42; Length 428;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-186;  
 Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 401 cagatggaggaagggccctgtaccatccaaagagccacgacagcttccacac 460  
 Db 396 CAGGATGGAGGAGGAGGCGCTGTACCCATCCAGAAGGCCACGACAGCTTCCACAC 337  
 QY 461 agaactccatcccggtggtcttctgattatgaagctgcacggcgaggtccacca 520  
 Db 336 AGAATCTCCATCCCGGGTGGCTTCTGGATCATTAAGCTGCCACGGGAGGTCCACCA 277  
 QY 521 ggaatccctggagggccactggtcagcgagaagcgacacgcctgcaggccatcg 580  
 Db 276 GGATGCCCTGGAGGGCGGCCACTGGCTCAGCGAGAAGGACACCGCTGCAGGCCATCG 217  
 QY 581 ggaatggactccgaaggggacccacaaagagctcctagaagaggggacgagctcctc 640  
 Db 216 GGATGGACTCCGCAAGGGGACCCCAAGGAGCTCCTAGAGGGGACCGAGAGTCTCTC 157  
 QY 641 ccactccaggtgtccccccaaagacccacttactgtacatctcctcagggcctctcgca 700  
 Db 156 CCACCTCCAGGCTGTCCCCCGGAGACCCACTTACTGTACATCTCAGCCCTCTCGGCA 97  
 QY 701 gctgtagggtgggacggggagcacctgctgtagcccccatcagacccctgccccaa 760  
 Db 96 GCTGTAGGGTGGGACCGGGAGCACCTGCTGTAGCCCCCATCAGACCCCTGCCCAAG 37

QY 761 caccataggaataaagtctttc 785  
 Db 36 CACCATATGAATAAAGTCTTTC 12

RESULT 5  
 AI139919/c  
 LOCUS  
 DEFINITION  
 3' similar to SW-IF4-CHICK Q90839 UNKNOWN LENS FIBER PROTEIN  
 CLFE5T4 PRECURSOR. ; mRNA sequence.  
 ACCESSION  
 AI139919  
 VERSION  
 AI139919.1  
 KEYWORDS  
 EST.  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 593)  
 AUTHORS  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL  
 Unpublished (1997)  
 COMMENT  
 On Jan 9, 1998 this sequence version replaced gi:936835.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone Distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1351 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham



FEATURES  
source  
High quality sequence stop: 461.  
Location/Qualifiers  
1..593

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1737911"  
/clone\_lib="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from Clontech  
Laboratories, Inc., and primed with a Not I - oligo(dT)  
primer [5'  
TGGTACCAATCGAAGTGGAGCGCGCCGCCAAATTTTTTTTTTTT 3']  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization to Cot5, and was  
constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 89 a 169 c 185 g 150 t  
ORIGIN

Query Match 48.5%; Score 381; DB 42; Length 593;  
Best Local Similarity 100.0%; Pred. No. 3e-184;  
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 atgaggagaagagagccctgtaccatccagaagggccacgagctccacagaa 464  
DB 392 ATGGAGGAGAGAGGCCCTGTACCCATCCAGAGGCCACGACAGCTTCCACAGAA 333  
QY 465 ctccatccccgggtggtcctcctggtatcattagctgccacggcgaggtccaccaggat 524  
DB 332 CTCATCCCGGGTGGCTTCTGGATCATTAAGCTGCCACGGGGAGGTCCACACAGAT 273  
QY 525 gccctgagggggccacgtgctcagcaggaagcagccgctgagggccatccggat 584  
DB 272 GCCTGGAGGGGGCCACTGGCTCAGCGAGAGCGACACCGCTGCAGGCCATCCGGGAT 213  
QY 585 ggactccgaaggaggaccacagagctcctagaagaggggaccagagctctccac 644  
DB 212 GGACTCCGAGAGGGACCCACAGAGCGTCTTAGAGAGGGGACCGAGAGCTCTCCAC 153  
QY 645 tcaggctgtcccccgaaagaccacttactgtacatcctcagggccctctggcagctg 704  
DB 152 TCAGGCTGTCCCGGAAAGACCCACTTACTGTACATCTCAGGCCCTCTCGCAGCTG 93  
QY 705 taggggtgggagggggagcaactcctgtagcccccatcagaccctgcccccaagcacc 764  
DB 92 TAGGGTGGGAGCCGGGAGCACTGCCTGTAGCCCCCATCAGACCCCTGCCCAAGCACC 33  
QY 765 atatggaataaagtctttc 785  
DB 32 ATATGGAATAAGTCTTTC 12

RESULT 6  
AI879701/c  
LOCUS  
DEFINITION AI879701 460 bp mRNA EST 23-AUG-1999  
au57c10.x1 schneider fetal brain 00004 Homo sapiens cDNA clone  
IMAGE:2518866 3', mRNA sequence.  
ACCESSION AI879701  
VERSION AI879701.1 GI:5553750  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 460)  
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,  
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,

TITLE  
JOURNAL  
COMMENT

Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
WashU-NCI human EST Project  
Unpublished (1997)  
On Feb 22, 1999 this sequence version replaced gi:4282756.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (infoimage.llnl.gov) for further information.  
Seq primer: 40UP from Gibco.  
Location/Qualifiers

FEATURES  
source

1..460  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2518866"  
/clone\_lib="Schneider fetal brain 00004"  
/sex="male"  
/tissue\_type="frontal lobe"  
/dev\_stage="5 months post-conception"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescript SK (Stratagene);  
Site\_1: SstI; Site\_2: XhoI; Double-stranded cDNA was  
prepared from human fetal brain tissue. 5' and 3'  
adaptors were used in cloning as follows: 5' adaptor  
sequence:  
5'-GAGAGAGAGAGAGCTCAAGGATCTTAATAAATTAATCCCCCCCCCCC-3'  
and 3' adaptor sequence:  
5'-GAGAGAGAGAGACTCGAGTATTTTTTTTTTTT-3'. The library was  
size-selected for >0.5 kb inserts and has an average  
insert size estimated at 1.2 kb. This library was  
constructed using the CAP-trapper method for full-length  
enrichment and has not undergone amplification. Library  
was constructed by Dr. Claudio Schneider (LNCIB-Area  
Science Park, Trieste, Italy)."  
BASE COUNT 73 a 133 c 142 g 111 t  
ORIGIN

Query Match 46.1%; Score 362; DB 62; Length 460;  
Best Local Similarity 100.0%; Pred. No. 1.6e-174;  
Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 tgttaccatccagaagggccacggacagcttcacacagaactccatccccgggtggcct 483  
DB 378 TGTATCCCATCCAGAGGCCACGACAGCTTCCACACAGACTCCATCCCGGTGGCCT 319  
QY 484 tctgatacttaagctgccacggcgagggtccacacagatgccctggaggcgccact 543  
DB 318 TCTGGATCATTAAGCTGCCACGGGAGGTCCACAGATGCCCTGAGGGCGGCACT 259  
QY 544 ggctcagcagaagcagacacccctgcagccatccgagatggactccgaggggaccc 603  
DB 258 GGCTCAGCAGAAGGCACACCCCTTCAGGCCATCCGGGTGAGCTCCCAAGGGGACCC 199  
QY 604 acaagcagctcctagaagggggacccagagctctccacactccaggtgtcccccggaa 663  
DB 198 ACAAGGACGCTCTAGAGAGGGGACCGAGAGCTCTCCACACTCCAGGCTGTCCCCCGAA 139  
QY 664 agaccacactactgtacatcctcagggccctctcggcagctgtagggtgggagccggga 723  
DB 138 AGACCCACTTACTGTACATCTCAGGCCCTCTCGGAGCTGTAGGGGTGGGACCGGGA 79  
QY 724 gacctgccttagcccccatcagaccctgccccagcaccataggaataaagtctt 783  
DB 78 GCACCTGCCTGTAGCCCCCATCAGACCCCTGCCCAAGCACCATATGGAATAAGTCTT 19  
QY 784 tc 785  
DB 18 TC 17

```

RESULT 7
AI377443/c 368 bp mRNA EST 18-MAR-1999
LOCUS tc23f04.x1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
DEFINITION IMAGE:2065471 3', mRNA sequence.
ACCESSION AI377443
VERSION AI377443
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 368)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT On Jun 15, 1998 this sequence version replaced gi:3226479.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 772 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 312.
FEATURES
Location/Qualifiers
source 1..368
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2065471"
/clone_lib="Soares_total_fetus_Nb2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGGCGCTTAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 56 a 97 c 123 g 92 t
ORIGIN
Query Match 43.9%; Score 345; DB 45; Length 368;
Best Local Similarity 100.0%; Pred. No. 7.7e-166;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 441 gccacgacagcttcacacagaaactccatcccggtggcctcttgatcaatgaagt 500
Db GCCACGGACAGCTCCACACAGAACTCCATCCCGGGTGGCTTCTGATCAITTAAGCTG 309

QY 501 ccacggcggaggtccacacagagatgcctggagggcgccactggtccagcgaagcga 560
Db CCACGGCGGAGGTCCACACAGAGATGCGCTGGAGGGCGGCCACTGGCTCAGCGAAGCGA 249

QY 561 caccgctgcagggcctcccggtgagctccgcaaggcgaccacaaaggagcgtcctgaa 620
Db CACCGCTGCAGGCGCATCCGGGATGGACTCCGAAGGGGACCCACAAAGAGCGTCTAGAA 189

QY 621 gagggaccgagagctcctccactccagggtgtcccccgaagaccactactgtac 680
Db GAGGGGACCGAGAGCTCTCCACTCCAGGCTGTCCCGCGAAGACCCACTACTGTAC 129

QY 681 atcctagccctctcgagcttaggggtgggagccggggagcacctccttagccc 740
Db ATCTCAGGCCCTCTCGCAGCTGTAGGGGTGGGGACCGGGGAGCACCTGCTGTAGCCC 69

```

```

QY 741 ccatcagaccctgccccaaagaccatataatgaaataaagtctttc 785
Db 68 CCATCAGACCCTGCCCAAGCACCACCATATGGAATAAAGTCTTTC 24

RESULT 8
AI200868/c 428 bp mRNA EST 14-OCT-1998
LOCUS qf62g12.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1754662
DEFINITION 3', mRNA sequence.
ACCESSION AI200868
VERSION AI200868
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 428)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
On Jan 14, 1998 this sequence version replaced gi:1877571.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbrrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 406.
FEATURES
Location/Qualifiers
source 1..428
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1754662"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5'
TGTTACCAATCTGAAGTGGAGCGGCGGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 67 a 125 c 136 g 100 t
ORIGIN
Query Match 43.6%; Score 342; DB 43; Length 428;
Best Local Similarity 99.7%; Pred. No. 2.6e-164;
Matches 392; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 393 aaggtaccacaggtgagagagaggccctggtaccatccagagccagcagc 452
Db AAGGTACCCAGGATGGAGGAGAGAGGAGGCCCTGGTCCCATCCAGAACGCCAGGACAGC 345

QY 453 ttccacagaactccatcccggtggcctcttgatcattaaagctccagcgcgagg 512
Db TTCCACAGAACTCCATCCCGGGTGGCTTCTGTGATCATTAAAGCTGCACGGCGAAGG 285

QY 513 tcccacagagatccctggaggcgccactggtcagcagagaagcagcacccctgcag 572

```

```

Db 284 TCCACAGGATGCCTGGAGGGCGCCACTGGCTCAGCGAAGGACCGCCCTGCAG 225
QY 573 gccatccggatggactccgcaaggggaccacaaagagcgtcctagaagaggggaccgag 632
Db 224 GCATCCGGGATGGATCCGCAAGGGGACCCACAAGGAGCTCTAGAAGAGGGGACCGAG 165
QY 633 agctctccactccaggtgtcccccgaagaccacttactgtacatctcagggcc 692
Db 164 AGCTCTCCCACTCAGGCTGTCCCCCGAAGACCCACTTACTGTATCTCTCAGGCC 105
QY 693 tctcgagctaggggtgggagccgggagcacctgctgtagcccccacacagacct 752
Db 104 TCTGGCAGCTAGGGGTGGGACCGGGAGCACCTGCTGTAGCCCCCATCAGACCT 45
QY 753 gcccaagcaccatattggaataaagtctttc 785
Db 44 GCCCAAGCACCATATGGAATAAAGTTCTTTC 12

RESULT 9
A123276/c
LOCUS A123276 443 bp mRNA EST 29-NOV-1998
DEFINITION q953el2.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1838926
3', mRNA sequence.
ACCESSION A123276
VERSION A123276.1 GI:3805479
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 443)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797691.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 646 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 439.
Location/Qualifiers
1. .443
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1838926"
/cdb_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5]
TGTACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 71 a 119 c 149 g 104 t
ORIGIN

```

```

Query Match 43.2%; Score 339; DB 43; Length 443;
Best Local Similarity 99.7%; Pred. No. 9e-163;
Matches 389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 396 gtaccacagatggagagaagagccctgtaccatccagaagccacgacagcttc 455
Db 401 GTACCAGATGGAGAGAAGAGGCCCTGTGATCATTAACTGCCAGGGGAGGTCC 342
QY 456 cacacagaactccatccccgggtgcttctggatcattaaagtcacacggcgaggtcc 515
Db 341 CACACAGAACTCATCCCGGGTGCCTTCTGGATCATTAACTGCCAGGGGAGGTCC 282
QY 516 caccaggatgcctggaggggccactggctcagcgaaagcagaccgcctcagggcc 575
Db 281 CACCAGGATGCCCTGGAGGCGCACCTGGCTCAGCGAAGCGACACCGCCTGCAGGC 222
QY 576 atccgggatgactccgaaggggaccacaaagagcgtcctagaagaggagaccagac 635
Db 221 ATCCGGGATGGACTCCGAAGGGGACCCACAAGAGCGTCTAGAAAGGGGACCGAGAC 162
QY 636 tctccactccagctgtccccccgaagaccacttactgtacatctcctcagggcctct 695
Db 161 TCTCCCACTCCAGGCTGTCCCCCGAAGACCCACTTACTGTATCTCTCAGGCCCTC 102
QY 696 cggcagctgtagggtgggagccgggagcacctgcctgtagcccccatcagaccctgcc 755
Db 101 CGGCAGCTGTAGGGTGGGACCGGGGAGCACCTGCTGTAGCCCCCATCAGACCTGCC 42
QY 756 ccaacacacatattggaataaagtctttc 785
Db 41 CCAAGCACCATATGGAATAAAGTTCTTTC 12

RESULT 10
A1138943/c
LOCUS A1138943 481 bp mRNA EST 28-OCT-1998
DEFINITION qd79a06.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1735666
3', mRNA sequence.
ACCESSION A1138943
VERSION A1138943.1 GI:3644915
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 481)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2286285.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 979 Std Error: 0.00
Seq primer: -40ml3 fwd ET from Amersham
High quality sequence stop: 456.
Location/Qualifiers
1. .481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1735666"
/cdb_lib="Soares_testis_NHT"
/sex="male"

```

FEATURES  
source

/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from Clontech  
Laboratories, Inc., and primed with a Not I - oligo(dT)  
primer [5']  
TGTACCATCTGAATGGAGGGCGCGCCCAATTTTTTTTTTTT 3']  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization to Cot5, and was  
constructed by Bento Soares and M. Fatima Bonaldo.  
BASE COUNT 77 a 142 c 147 g 113 t 2 others  
ORIGIN

Query Match 40.0%; Score 314; DB 42; Length 481;  
Best Local Similarity 100.0%; Pred. No. 5.5e-150;  
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 472 ccgagggtgcttgcattgaagtcgacgagggaggtccaccagatgcctgg 531  
Db 328 CCCGGTGGCTTCTGGATCATTAAGTCGCACGGGAGGTCCACCAGATGCCCTGG 269  
QY 532 agggcgccactgctcagcgagaagcgacacgcctgcaggccatccgggatgactcc 591  
Db 268 AGGCGGCCACTGCTCAGCGAGAAGCGACACCGCCTGCAGGCCATCCGGGATGGACTCC 209  
QY 592 qcaaggggaccacagagctctagaaggggagggagagctctccactccagggc 651  
Db 208 GCAAGGGACCCACAGAGAGCTCTAGAAGAGGGGACCGAGAGTCTCTCCACTCCAGGC 149  
QY 652 tgtccccccaaagaccacttactgtacatctcagccctctcggcagctgtagggt 711  
Db 148 TGTCCCCCGAAGACCCACTTACTGTACATCTCTCAGGCCCTCTCGCAGCTGTAGGGT 89  
QY 712 gggagcggggagacactgctgtagcccccatcagaccctgccccagaccatatgga 771  
Db 88 GGGAGCCGGGAGACCTGCTGTAGCCCCCATCAGACCCTGCCCAAGCACCATATGGA 29  
QY 772 aataaagtcttttc 785  
Db 28 AATAAGTCTTTC 15

RESULT 11  
AI652314/c 382 bp mRNA EST 04-MAY-1999  
LOCUS wb60a03.x1 NCI\_CGAP\_GC6 Homo sapiens cDNA clone IMAGE:2310028 3',  
DEFINITION mRNA sequence.  
ACCESSION AI652314  
VERSION AI652314.1 GI:4736293  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 382)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
Tumor Gene Index  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Unpublished (1997)  
On May 7, 1998 this sequence version replaced gi:3121068.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/dbrrp/Image/Image.html  
Seq primer: -40UP from Gibco.  
FEATURES  
source Location/Qualifiers  
1.382  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2310028"  
/clone\_lib="NCI\_CGAP\_GC6"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Plasmid DNA from the normalized library  
NCI-CGAP\_GC4 was prepared, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from a pool of 5,000 clones made  
from the same library (clones 1257096-1258631,  
1469064-1470983, and 1475592-1476743). Subtraction by  
Bento Soares and M. Fatima Bonaldo.  
BASE COUNT 59 a 103 c 131 g 89 t  
ORIGIN

Query Match 34.1%; Score 268; DB 49; Length 382;  
Best Local Similarity 99.5%; Pred. No. 1.8e-126;  
Matches 368; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 416 gggggcctggtaccatccagaagggcagcgagcttccacagaactccatcccg 475  
Db 382 GGAGGGCCCTGGTACCCATCCAGAAGGCCACGACAGCTTCCACACAGAACTCCATCCCG 323  
QY 476 ggtggccttctgattcattgaagtcgacggaggtccacagagtcctcctgagg 535  
Db 322 GGTGGCTTCTGTGATCATTAAGCTGCCAGGGAGGTCCACAGATGCTCCCTGGAGG 263  
QY 536 cggccactggtcagcgagaagcgacacgcctgcaggccatccggatggactccgcaa 595  
Db 262 CAGCCACTGGCTCAGCGAGAAGCGACACCGCTGCAGGCCATCCGGGATGGACTCCGCAA 203  
QY 596 ggggaccacagagctcctagaaggggagggagagctcctccactccaggtgtc 655  
Db 202 GGGAGCCACAGAGAGCTCTTAAAGAGGGGACCGAGAGCTCTCCACTCCAGGTGTC 143  
QY 656 ccccgaaagaccacttactgtacatctcagccctctcggcagctgtagggtgggg 715  
Db 142 CCCCCGAAAGACCCACTTACTGTACATCTCTAGGCCCTCTCGGCAGCTGTAGGGGTGGG 83  
QY 716 accggggagacactgctgtagcccccatcagaccctgccccagaccatatggaata 775  
Db 82 ACCGGGAGCACCTGCTGTAGCCCCCATCAGACCCTGCCCAAGCACCATATGGAATA 23  
QY 776 agttcttctc 785  
Db 22 AAGTCTTCTTC 13

RESULT 12  
AA854987/c 269 bp mRNA EST 31-DEC-1998  
LOCUS aj53g06.s1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:1394074  
DEFINITION 3', mRNA sequence.  
ACCESSION AA854987  
VERSION AA854987.1 GI:2942525  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 269)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
Tumor Gene Index  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL  
COMMENT  
Unpublished (1997)  
On Jan 17, 1998 this sequence version replaced gi:1900909.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 956 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 214.

FEATURES  
source  
1..269  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1394074"  
/clone\_lib="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo (dT) primer [5', TGTTACCAATCTGAAGTGGGAGCGCGCCCAATTTTTTTTTTTT 3'].  
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 42 a 74 c 91 g 62 t

Query Match 32.9%; Score 258; DB 39; Length 269;  
Best Local Similarity 100.0%; Pred. No. 2.3e-121;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 ctggaggcgccgactggctcagcgagcgacacccgctgcagggccatccgggatgga 587  
Db CTGGAGGGGGCCACCTGGCTAGCGAGAAAGCGACACCGCTGCAGGCCATCCGGGATGA 210

QY 588 ctccgcaaggacccacacagacgtccctagaagaggggacggagagctccctccactcc 647  
Db CTCCGCAAGGGGACCCACAGGACGTCTAGAGAGGGGACCGAGAGCTCCTCCCACTCC 150

QY 648 aggtgtctcccccgaaagaccacttactgtacatctcctcaggccctctcggcagctgtg 707  
Db AGGCTGTCCCCCGAAGACCCACTTACTGTACATCTTCAGGCCCTCTCGGCAGCTGTAG 90

QY 708 ggggtgggacccggggagaccctcgtgtgcccccatagaccctccccaagaccata 767  
Db GGGTGGGGACCGGGAGACCTCTGCTGTAGCCCCCATCAGACCCCTGCCCAAGCACCAT 30

QY 768 tggaataaagtctttc 785  
Db TGGAAATAAAGTCTTTTC 12

RESULT 13  
AI004529/c  
LOCUS AI004529 257 bp mRNA EST 27-AUG-1998  
DEFINITION ot57g08.s1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:1620926  
3', mRNA sequence.  
ACCESSION AI004529  
VERSION AI004529.1 GI:3214039  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 257)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Feb 11, 1998 this sequence version replaced gi:2339714.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 965 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 220.

FEATURES  
source  
1..257  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1620926"  
/clone\_lib="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo (dT) primer [5', TGTTACCAATCTGAAGTGGGAGCGCGCCCAATTTTTTTTTTTT 3'].  
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 41 a 67 c 88 g 61 t

Query Match 31.3%; Score 246; DB 41; Length 257;  
Best Local Similarity 100.0%; Pred. No. 3.1e-115;  
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 cactggctcagcagaagcgacacccgctgcagggccatccgggatggactccgcaagggg 599  
Db 257 CACTGGCTCAGCAGAAGCGACACCGCTGCAGGCCATCCGGATGTGACTCCGCAAGGG 198

QY 600 acccacaaggagctcctagaagaggggacggagagctcctccactccaggtgtcccc 659  
Db 197 ACCCACAAGGAGCTCCTAGAAGAGGGGACCGAGAGCTCCTCCACTCCAGGTGTCCCC 138

QY 660 cgaagaccacttactgtacatctcctcaggccctcctggcagctgtagggtgggaccg 719  
Db 137 CGAAAGACCCACTTACTGTACATCTCTAGGCCCTCTCGGCAGCTGTAGGGGGGGACCG 78

QY 720 gggagcacctgctgtagcccccatcagaccctgcccccaagcaccatatggaaataaagt 779  
Db 77 GGGAGCACTGCTGTAGCCCCCATCAGACCTGCCCCAAGCACCACCATATGGAATAAAGT 18

QY 780 tctttc 785  
Db 17 TCTTTC 12

RESULT 14

```

AA397836      440 bp      mRNA      EST      16-MAY-1997
LOCUS      zt77h04.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:728407
DEFINITION      5' similar to TR:G517093 G517093 HYPOTHETICAL 39.2 KD PROTEIN. ;,
                mRNA sequence.
ACCESSION      AA397836
VERSION        AA397836.1 GI:2050593
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 440)
AUTHORS        Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
                Kucaba,T., Lacy,M., Lennon,G., Marra,M., Martin,J.,
                Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
                White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE          Washu-Merck EST Project 1997
JOURNAL        Unpublished (1997)
COMMENT        On Dec 30, 1996 this sequence version replaced gi:1528967.
                Contact: Wilson RK
                Washington University School of Medicine
                4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
                Tel: 314 286 1800
                Fax: 314 286 1810
                Email: est@watson.wustl.edu
                This clone is available royalty-free through LNL; contact the
                IMAGE Consortium (info@image.llnl.gov) for further information.
                Possible reversed clone: similarity on wrong strand
                Seq primer: -28ml3 rev2 ET from Amersham
                High quality sequence stop: 392.
FEATURES       Location/Qualifiers
                source
                1..440
                /organism="Homo sapiens"
                /db_xref="GDB:5925320"
                /db_xref="taxon:9606"
                /clone="IMAGE:728407"
                /clone_lib="Soares_testis_NHT"
                /sex="male"
                /lab_host="DH10B"
                /notes="Vector: pT73D-Pac (Pharmacia) with a modified
                polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                was prepared from mRNA obtained from Clontech
                Laboratories, Inc., and primed with a Not I - oligo(dT)
                primer [5'.
                TGTACCAATCTGAAGTGGGAGCGCGCCCAATTTTTTTTTTTT 3'].
                Double-stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia), digested with Not I and cloned into the Not I
                and Eco RI sites of the modified pT73 vector. Library
                went through one round of normalization to Cot5, and was
                constructed by Bento Soares and M. Fatima Bonaldo. "
                BASE COUNT      98 a 142 c 120 g 80 t
ORIGIN
Query Match      19.5%; Score 153; DB 33; Length 440;
Best Local Similarity 100.0%; Pred. NO. 1.1e-67;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 cttttcgtaaagttaacctgttcgggcatacacagcttattctgtcccccattgac 302
Db 219 CTTTCTCTGAANGTAACTCTCTGGGGCATACAGACTATTCTCTGCCCCCATGGAC 278

QY 303 ttccggggccctccctgggaactaccacaaagagagagacagagaccagctggggaac 362
Db 279 TTCCGGGGCCCTCTGGGAATACCACAAAGAGAGAGAACAGGAGACACCACTGGGGAAC 338

QY 363 aacacctctccagccacctccagatcgacaag 395
Db 339 AACACCTCTCCAGCCACCTCCAGATCGACAAG 371

RESULT 15

```

```

AI879324      528 bp      mRNA      EST      23-AUG-1999
LOCUS      au57c10.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone
DEFINITION      IMAGE:2518866 5', mRNA sequence.
ACCESSION      AI879324
VERSION        AI879324.1 GI:5553373
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 528)
AUTHORS        Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
                Krizman,D., Kucaba,T., Lacy,M., Lennon,G., Marra,M.,
                Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
                Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE          WashU-NCI human EST Project
JOURNAL        Unpublished (1997)
COMMENT        On Jun 22, 1998 this sequence version replaced gi:3247250.
                Other ESTs: au57c10.x1
                Contact: Wilson RK
                Washington University School of Medicine
                4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
                Tel: 314 286 1800
                Fax: 314 286 1810
                Email: est@watson.wustl.edu
                This clone is available royalty-free through LNL; contact the
                IMAGE Consortium (info@image.llnl.gov) for further information.
                Seq primer: -40RP from Gibco
                High quality sequence stop: 392.
FEATURES       Location/Qualifiers
                source
                1..528
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:2518866"
                /clone_lib="Schneider fetal brain 00004"
                /sex="male"
                /tissue_type="frontal lobe"
                /dev_stage="5 months post-conception"
                /lab_host="DH10B"
                /notes="Organ: brain; Vector: pBluescript SK (Stratagene);
                Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
                prepared from human fetal brain tissue. 5' and 3'
                adaptors were used in cloning as follows: 5' adaptor
                sequence:
                5'-GAGAGAGAGAGAGAGCTCAAGATCCTTAATTAATTAATCCCGCCCCCCCC-3'
                and 3' adaptor sequence:
                5'-GAGAGAGAGAGAGCTCGAGTGTGTGTGTGTGTGTGTGTGTGT-3'. The library was
                size-selected for >0.5 kb inserts and has an average
                insert size estimated at 1.2 kb. This library was
                constructed using the CAP-trapper method for full-length
                enrichment and has not undergone amplification. Library
                was constructed by Dr. Claudio Schneider (LNCIB-Area
                Science Park, Trieste, Italy)".
                BASE COUNT      101 a 144 c 186 g
ORIGIN
Query Match      18.2%; Score 143; DB 62; Length 528;
Best Local Similarity 100.0%; Pred. NO. 1.5e-62;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cggaaatccggaggtccgggtgacccgggctgtgtctagcataaaggcgagccagaaga 60
Db 335 CGGAATCCGGAGTCCGGTGACCCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 394

QY 61 aggggggggtatgggagaagctccccacctccccccgcaaggcgcatctctgtgtcc 120
Db 395 AGGGGGGGGTATGGGAGAAGCTCCCCACCTCCCCCGGCAAGCGCATCTGTGTGTCT 454

QY 121 tgctgtgtctctctctaccctg 143
Db 455 TGCTGTGTCTCTCTCTACCTGT 477

```

Search completed: May 1, 2000, 14:44:51  
Job time: 14011 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:47:50 ; Search time 4425.31 Seconds  
(without alignments)  
-585.278 Million cell updates/sec

Title: US-09-215-435-123  
Perfect score: 853  
Sequence: 1 ggaggatggcgagcagtct.....tttgcgaaaaaaaaaaaaa 853

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0  
Searched: 821193 seqs, -1518192014 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0  
Maximum DB seq length: 1000000  
Post-processing: Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_v1:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl1:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_v1:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*

44: gb\_htg6:\*  
45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	469	55.0	1998	10	HSM800599	AF080097 Homo sapi
2	22	2.6	9384	8	AF076243	AF076243 Arabidops
C 3	22	2.6	172650	44	AC010993	AC010993 Drosophil
C 4	22	2.6	201170	41	AC009788	AC009788 Homo sapi
C 5	21	2.5	311	7	RICKN6922	D17786 Rice mRNA f
C 6	21	2.5	858	8	AF191099	AF191099 Fagopyrum
7	21	2.5	1390	7	SOALDCYT	X65742 S.oleracae
8	21	2.5	1493	8	HSM80598	U80598 Hevea bras
9	21	2.5	2424	10	HSM801097	AL117570 Homo sapi
C 10	21	2.5	26894	35	CELH43107	AF125965 Caenorhab
C 11	21	2.5	67220	45	AC017014.3	Continuation (4 of
C 12	21	2.5	110000	45	AC017014.2	Continuation (3 of
C 13	21	2.5	291288	42	AC008878	AC008878 Homo sapi
14	20	2.3	297	8	AF047051	AF047051 Glycine m
15	20	2.3	949	34	CEU82968	U82968 Caenorhabdi
16	20	2.3	964	7	ATRA82LPR	Y09314 A.thaliana
17	20	2.3	986	8	AF165529	AF165529 Rumex pal
18	20	2.3	1247	8	AF139818	AF139818 Brassica
C 19	20	2.3	1419	5	A80078	A80078 Sequence 1
C 20	20	2.3	1419	5	AR029597	AR029597 Sequence
C 21	20	2.3	1419	5	AR049459	AR049459 Sequence
C 22	20	2.3	1419	5	AR066452	AR066452 Sequence
C 23	20	2.3	1419	5	I16769	I16769 Sequence 1
24	20	2.3	2242	7	YSCF395C	M80922 Yeast trans
C 25	20	2.3	2641	7	SCYBR123C	Z35922 S.cerevisia
26	20	2.3	3004	7	YSCF1C1	M63385 S.cerevisia
27	20	2.3	4437	5	AR049083	AR049083 Sequence
28	20	2.3	4437	10	HSU39817	U39817 Human Bloom
29	20	2.3	40763	45	AC017458	AC017458 Drosophil
C 30	20	2.3	69748	7	SCRAC11	X78993 S.cerevisia
C 31	20	2.3	79097	33	AC006897	AC006897 Caenorhab
C 32	20	2.3	81004	8	ATAC002521	AC002521 Arabidops
C 33	20	2.3	115382	42	AC010689	AC010689 Drosophil
C 34	20	2.3	127933	41	AC008849	AC008849 Homo sapi
C 35	20	2.3	171772	44	AC009584	AC009584 Homo sapi
C 36	20	2.3	186331	40	AC007376	AC007376 Homo sapi
37	20	2.3	233904	44	AC012517	AC012517 Homo sapi
38	20	2.3	298216	33	AC006875	AC006875 Caenorhab
39	19	2.2	1031	5	AR016872	AR016872 Sequence
40	19	2.2	1031	5	AR020898	AR020898 Sequence
41	19	2.2	1031	5	AR027221	AR027221 Sequence
42	19	2.2	1031	5	AR038508	AR038508 Sequence
43	19	2.2	1031	5	AR064650	AR064650 Sequence
44	19	2.2	1031	5	AR067575	AR067575 Sequence
45	19	2.2	1031	5	I38527	I38527 Sequence 10

ALIGNMENTS

RESULT 1  
HSM800599  
LOCUS  
DEFINITION  
ACCESSION  
VERSION

HSM800599 1998 bp mRNA PRI 23-JUN-1999  
Homo sapiens mRNA; cDNA DKFZp564P0462 (from clone DKFZp564P0462).  
AL080097  
AL080097.1 GI:5262519

```

KEYWORDS      human.
SOURCE        Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 1998)
AUTHORS       Blum.H., Bauersachs.S., Mewes.H.W., Gassenhuber.J. and Wiemann.S.
TITLE         Direct Submission
JOURNAL       Submitted (22-JUN-1999) MIPS, Am Klopferspitze 18a D-82152
              Martinsried, GERMANY
COMMENT       Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
              Research Center (DKFZ); sequenced by LMU within the cDNA sequencing
              consortium of the German Genome Project. This clone is available at
              the RZPD in Berlin.
              Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
              Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES     Location/Qualifiers
              source
                1..1998
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone_lib="564 (synonym: hfbr2). Vector pAMPI; host
                    Xl-2blue; sites NotI + SalI"
                  /dev_stage="fetal"
                  /tissue_type="brain"
                  /note="corresponding STS: EMBL:HS640145"
              gene
                1..802
                  /gene="DKFZp564P0462"
              CDS
                <1..802
                  /gene="DKFZp564P0462"
                  /note="unknown"
                  /codon_start=2
                  /product="hypothetical protein"
                  /protein_id="CAB45709.1"
                  /db_xref="GI:5262520"
                  /translation="GGAGGWSLSNARMNRRTAFVIACVLSLSTIYMAASIGTDF
                    WVEYRSPVQENSDEKSIWDEFISDEADEKYNDALEFNGTVGLWRRCIPIPKNMH
                    WSPPTERTSFVVTVCVFTLEQWMEFVDPGNHNSGIDLLRTYLRCQELLFPVS
                    LGLMGFALIGLACICRSLYPTIATGLHLAGLCTLGSCVYAGIELLHOKLELP
                    DNVSFGFSGFLACVSPLOFMASALFIWAHTNRKEYTLMKAYRVA"
              polyA_signal
                1961..1966
              polyA_site
                595 a 379 c 357 g 667 t
              BASE COUNT
              ORIGIN
                Query Match 55.0%; Score 469; DB 10; Length 1998;
                Best Local Similarity 99.6%; Pred. No. 5.7e-247;
                Matches 569; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
              QY 1 ggagatggcgagcagctcgaatgcagatggataacccgttttgcacagcatttga 60
                |||||
              Db 11 ggagatggcgagcagctcgaatgcagatggataacccgttttgcacagcatttga 70
                |||||
              QY 61 attcctgtgtgcttagcctcattccaccatctcacatggcagcctcattggcacagac 120
                |||||
              Db 71 attcctgtgtgcttagcctcattccaccatctcacatggcagcctcattggcacagac 130
                |||||
              QY 121 ttctggtatgaatcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcga 180
                |||||
              Db 131 ttctggtatgaatcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcga 190
                |||||
              QY 181 tggatgaattcattatgtagatgaagcagatgaagcagatgaagcagatgaagcagatga 240
                |||||
              Db 191 tggatgaattcattatgtagatgaagcagatgaagcagatgaagcagatgaagcagatga 250
                |||||
              QY 241 tacaatggcagatgggattgtgagacggtgtatcaccatacccaaaacacatcattgg 300
                |||||
              Db 251 TACAATGGCAGATGGGATTGTGGAGACGGTGTATCACCATACCCAAAACATCATTGG 310
                |||||
              QY 301 tatagccaccagaagagacagatcatttgcattgtgtgtgcacaaaatgtgtgatttcaca 360
                |||||
              Db 311 TATAGCCACCAGAAAGGACAGAGTCATTTGATGTGGTCACAAAATGTGTGAGTTTCACA 370
                |||||
              QY 361 ctaactgagcagcttcattgagaaatttgcacccggaacacacataagcggattgat 420
                |||||
              Db 371 CTAACGTGAGCAGCTTCATGAGAAATTTGTTGATCCGGAACACACAATAGCGGATTGAT 430
                |||||
              QY 421 ctcttagaacctattcttggcgttgcagcttccttacccttttgcagtttaggtttg 480
                |||||
              Db 431 CTCCTTAGACCACTATCTTTGGCGTTGCCAGTTCCTTTTACCTTTTGTGAGTTAGGTTTG 490
                |||||
              QY 481 atgtcttggcgttgcagcttcattgagaaatttgcacccggaacacacataagcggattgat 540
                |||||
              Db 491 ATGTGTTTGGGCTTTGATCAGACTTTGCTTTCATTTGCCGAGCTTATATCCACC 550
                |||||
              QY 541 attgcacggcgatttcctcatctccttgcag 571
                |||||
              Db 551 ATGCCACGGGCATCTCCCATCTCCTTGCAG 581
                |||||
              RESULT 2
              AF076243 99384 bp DNA PLN 11-MAY-1999
              LOCUS Arabidopsis thaliana BAC T26N6 from chromosome IV at 19.3 CM,
              DEFINITION complete sequence.
              ACCESSION AF076243
              VERSION AF076243.1 GI:3309276
              KEYWORDS HTG.
              SOURCE Arabidopsis thaliana
              ORGANISM Arabidopsis thaliana
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
              eudicots; Rosidae; eurosid II; Brassicales; Brassicaceae;
              Arabidopsids.
              REFERENCE 1 (bases 1 to 99384)
              AUTHORS Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.
              TITLE Genomic sequence of Arabidopsis thaliana BAC T26N6, chromosome IV,
              19.3 CM
              JOURNAL Unpublished
              REFERENCE 2 (bases 1 to 99384)
              AUTHORS Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.
              TITLE Direct Submission
              JOURNAL Submitted (02-JUL-1998) Applied Biosystems, Division of Perkin
              Elmer, 850 Lincoln Center Drive, Foster City, CA 94404, USA
              REFERENCE 3 (bases 1 to 99384)
              AUTHORS Parnell,L.D.
              TITLE Direct Submission
              JOURNAL Submitted (11-MAY-1999) Lita Annenberg Hazen Genome Sequencing
              Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
              Harbor, NY 11724, USA
              REMARK Arabidopsis thaliana BAC T26N6 from chromosome IV at 19.3 CM
              COMMENT BAC T26N6 was sequenced as part of the Arabidopsis genome
              sequencing effort of the Cold Spring Harbor Consortium. For
              additional information, please see http://www.cshl.org/Arabweb.
              T26N6 is known to carry the m306 marker. Fingerprint data indicate
              that T26N6 overlaps with T19B17 and F4H6 of YAC C1C3F1.
              FEATURES
                Location/Qualifiers
                  1..99384
                    /organism="Arabidopsis thaliana"
                    /cultivar="Columbia"
                    /db_xref="taxon:3702"
                    /chromosome="IV"
                    /map="19.3 CM"
                    /clone="BAC T26N6"
                    /note="hybridizing YAC is C1C8B1"
                  1..9514
                    /note="overlap with T19B17, GenBank accession number
                      AF069441, from position 1 to 9513; there is a 1-bp
                      discrepancy between T26N6 and T19B17"
                  4268..4587
                    /note="function-unclassified; similar to K23L20, GenBank
                      accession number AB018874; similar to T19K24, GenBank
                      accession number AC002342"
                  5477..6145
                    /rpt_type="dispersed"
                    gene

```



```
/translation="MPSDNTFVRVRRQQRDPLHVNPSVGGNISVGSGRSRSGSTVP
Query Match          2.6%; Score 22; DB 8; Length 99384;
Best Local Similarity 100.0%; Pred. NO. 0.32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 832 gttttcgaaataaaaaa 853
Db 5668 GTTTTCGAAATAAAAAA 56689

RESULT 3
AC010993/c
LOCUS
DEFINITION
AC010993 172650 bp DNA HTG 22-NOV-1999
Drosophila melanogaster chromosome X clone BACR27005 (D877) RPCI-98
27.0.5 map 11F-12A strain y; cn bw sp. *** SEQUENCING IN PROGRESS
***, 122 unordered pieces.

AC010993
VERSION AC010993.6 GI:6460999
KEYWORDS HTG: HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 172650)
Celiker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Svirskaas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 172650)
Celiker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Pfeiffer,B., Poan,L., Sequira,A., Sethi,H., Snir,E.,
Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (29-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Nov 22, 1999 this sequence version replaced gi:6453830.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 122 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
663: contig of 663 bp in length
664 743: gap of unknown length
744 1199: contig of 456 bp in length
1200 1279: gap of unknown length
1280 1493: contig of 214 bp in length
1494 1573: gap of unknown length
1574 1959: contig of 386 bp in length
1960 2039: gap of unknown length
2040 2473: contig of 434 bp in length
2474 2553: gap of unknown length

2554 2995: contig of 442 bp in length
2996 3075: gap of unknown length
3076 3570: contig of 495 bp in length
3571 3650: gap of unknown length
3651 4246: contig of 596 bp in length
4247 4326: gap of unknown length
4327 4774: contig of 448 bp in length
4775 4855: gap of unknown length
4856 5416: contig of 561 bp in length
5417 5495: gap of unknown length
5496 5962: contig of 467 bp in length
5963 6042: gap of unknown length
6043 6523: contig of 481 bp in length
6524 7415: contig of 812 bp in length
7416 7495: gap of unknown length
7496 7914: contig of 419 bp in length
7915 7994: gap of unknown length
7995 8415: contig of 420 bp in length
8416 8494: gap of unknown length
8495 9009: contig of 515 bp in length
9010 9089: gap of unknown length
9090 9393: contig of 304 bp in length
9394 9474: gap of unknown length
10005: contig of 532 bp in length
10085: gap of unknown length
10793: contig of 708 bp in length
10874 11486: contig of 613 bp in length
11487 11566: gap of unknown length
11567 12486: contig of 920 bp in length
12487 13266: gap of unknown length
13267 13376: gap of unknown length
13377 14097: contig of 721 bp in length
14098 14177: gap of unknown length
14178 15086: contig of 909 bp in length
15087 15166: gap of unknown length
15167 16188: contig of 1022 bp in length
16189 16268: gap of unknown length
16269 16822: contig of 554 bp in length
16830 16902: gap of unknown length
16903 17518: contig of 616 bp in length
17519 17598: gap of unknown length
17599 17991: contig of 393 bp in length
17992 18071: gap of unknown length
18072 18926: contig of 855 bp in length
18927 19006: gap of unknown length
19007 19801: contig of 795 bp in length
19802 19881: gap of unknown length
19882 20938: contig of 977 bp in length
20939 22152: contig of 1214 bp in length
22153 22322: gap of unknown length
22323 23523: contig of 1291 bp in length
23524 23603: gap of unknown length
23604 25164: contig of 1561 bp in length
25165 25244: gap of unknown length
25245 26181: contig of 937 bp in length
26182 26261: gap of unknown length
26262 27020: contig of 759 bp in length
27021 27100: gap of unknown length
27101 28200: contig of 1100 bp in length
28201 30014: contig of 1734 bp in length
30015 30944: gap of unknown length
30945 31586: contig of 1492 bp in length
31587 31666: gap of unknown length
31667 32583: contig of 917 bp in length
32584 32664: gap of unknown length
32665 33514: contig of 851 bp in length
33515 33594: gap of unknown length
33595 34561: contig of 967 bp in length
```



\* \* 95926 115235: gap of unknown length  
 \* \* contig of 19310 bp in length  
 \* \* 115236 153747: contig of unknown length  
 \* \* contig of 38512 bp in length  
 \* \* 153748 185525: gap of unknown length  
 \* \* contig of 31778 bp in length  
 \* \* 185526 201170: contig of unknown length  
 \* \* contig of 15645 bp in length.

## FEATURES

source  
 1. .201170  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"  
 /clone="83\_F\_13"  
 /clone\_lib="RPCI-11 Human Male BAC"  
 BASE COUNT 61699 a 39980 c 39157 g 58910 t 1424 others  
 ORIGIN

Query Match 2.6%; Score 22; DB 41; Length 201170;  
 Best Local Similarity 100.0%; Pred. No. 0.3;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 454 ctttacctttgtgagtttag 475  
 |||||  
 Db 152177 CTTTACCTTTGTGAGTTAG 152156

## RESULT 5

RICKN6922/c 311 bp mRNA PLN 04-FEB-1999  
 LOCUS  
 DEFINITION Rice mRNA for ubiquitin-conjugating enzyme, partial sequence.  
 ACCESSION D17786  
 VERSION D17786.1 GI:455499  
 KEYWORDS ubiquitin-conjugating enzyme.  
 SOURCE Oryza sativa callus CDNA to mRNA.  
 ORGANISM Oryza sativa

REFERENCE  
 1 (bases 1 to 311)  
 Uchimiya,H.  
 Direct Submission  
 Submitted (28-SEP-1993) to the DDBJ/EMBL/GenBank databases.  
 Hirofumi Uchimiya, Institute of Mol. & Cell. Bioscience, The  
 University of Tokyo, Department of Cellular Function; 1-1-1 Yayoi,  
 Bunkyo-ku, Tokyo 113, Japan (E-mail:huchimiya@tansei.cc.u-tokyo.ac.j  
 p, Tel:03-3812-2111(ex.7844), Fax:03-3812-2910)  
 2 (bases 1 to 311)  
 Uchimiya,H.  
 Unpublished (1993)  
 Submitted (28-Sep-1993) to DDBJ by:  
 Hirofumi Uchimiya  
 Institute of Molecular and Cellular Biosciences  
 University of Tokyo  
 Bunkyo-ku, Tokyo 113  
 Japan  
 Phone: 03-3812-2111 x7844  
 Fax: 03-3812-2910.

## FEATURES

source  
 1. .311  
 /organism="Oryza sativa"  
 /db\_xref="taxon:4530"  
 /tissue\_type="callus"  
 1. .311  
 /gene="KN69"  
 <1..>311  
 /gene="KN69"  
 /codon\_start=1  
 /product="ubiquitin-conjugating enzyme"  
 /protein\_id="BAA21006.1"  
 /db\_xref="GI:4433363"

## gene

## CDS

/translation="FFFFFFRKRKAQATPLFVPINRDNMIQFPKPPKSFRTKVFHPNIN  
 SNGSICLDILKEQWSPALTIISKVLLSICLLTDPNGDDPLVPEIAHMYKTDRAKYEFT  
 AR"  
 BASE COUNT 83 a 80 c 61 g 87 t  
 ORIGIN

Query Match 2.5%; Score 21; DB 7; Length 311;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 833 ttttcgaaaaaataaaaaa 853  
 |||||  
 Db 22 TTTTGCAGAAAAAATAAAAA 2

## RESULT 6

AF191099 858 bp mRNA PLN 23-NOV-1999  
 LOCUS  
 DEFINITION Fagopyrum esculentum 1-Cys peroxiredoxin (Per1) mRNA, complete cds.  
 ACCESSION AF191099  
 VERSION AF191099.1 GI:6456095  
 KEYWORDS common buckwheat.  
 SOURCE Fagopyrum esculentum  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Eukaryota; Viridiplantae; Streptophyta; Magnoliophyta; eudicotyledons; core  
 eudicotyledons; Caryophyllales; Caryophyllales; Polygonaceae; Fagopyrum.  
 REFERENCE  
 1 (bases 1 to 858)  
 Lewis,M.L., Miki,K. and Ueda,T.  
 REFERE Ferri, a gene encoding an evolutionarily conserved 1-Cys  
 peroxiredoxin in buckwheat (Fagopyrum esculentum Moench), is  
 expressed in a seed-specific manner and induced during seed  
 imbibition  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 858)  
 Lewis,M.L., Miki,K., Muto,A. and Ueda,T.  
 AUTHORS Direct Submission  
 TITLE Submitted (30-SEP-1999) Bioscience, Salem-Teikyo University, 223  
 JOURNAL West Main Street, Salem, WV 26426-0500, USA

## FEATURES

source  
 1. .858  
 /organism="Fagopyrum esculentum"  
 /strain="Moench"  
 /db\_xref="taxon:3617"  
 1. .858  
 /genes="Per1"  
 44..703  
 /genes="Per1"  
 /codon\_start=1  
 /product="1-Cys peroxiredoxin"  
 /protein\_id="AAF12782.1"  
 /db\_xref="GI:6466096"  
 /translation="MPGLIGDSIPNLQVETTHGSKLHDFIGDSWILFSPHGDFTP  
 VCTTELKWKAKYEETKRGVLLGLSCDDIASHKWKIDVEAFPGSKVRYPIIADP  
 KREVTILKNVDPDSSGSLPSRALHVGPDKKVKSFLFLPATIGRNWEEVVRV  
 ESLQAANDKVKATPVDVQPGDEAVISPSVSDEEAKMFPHGTYFDLPSSKKGYLRFQT  
 V"  
 BASE COUNT 242 a 189 c 221 g 206 t  
 ORIGIN

Query Match 2.5%; Score 21; DB 8; Length 858;  
 Best Local Similarity 100.0%; Pred. No. 1.8;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 833 ttttcgaaaaaataaaaaa 853  
 |||||  
 Db 835 TTTTGCAGAAAAAATAAAAA 855

## RESULT 7

SOALDCYT

```
LOCUS       SOALDCYT      1390 bp      mRNA           PLN      30-JUN-1993
DEFINITION   S. oleraceae ALDCYT mRNA for fructose-1,6-bisphosphate aldolase.
ACCESSION    X65742.534156
VERSION      X65742.1 GI:22619
KEYWORDS     fructose-1,6-bisphosphate aldolase.
SOURCE       spinach.
ORGANISM     Spinacia oleracea
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;
              Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
REFERENCE    1 (bases 1 to 1390)
AUTHORS      Pelzer-Reith,B.
TITLE        Direct Submission
JOURNAL      Submitted (05-JUN-1992) B. Pelzer-Reith, Inst. f. Pflanzenphys. u.
              Mikrobiologie, FU Berlin, Konigin-Luise-Str. 12-16a, 1000 Berlin
              33, FRG
FEATURES     source
              2 (bases 1 to 1390)
              AUTHORS      Pelzer-Reith,B., Penger,A. and Schnarrenberger,C.
              TITLE        Plant aldolase: cDNA and deduced amino-acid sequences of the
              JOURNAL      chloroplast and cytosol enzyme from spinach
              MEDLINE      Plant Mol. Biol. 21 (2), 331-340 (1993)
              FEATURES     93144707
              Location/Qualifiers
              1..1390
              /organism="Spinacia oleracea"
              /db_xref="taxon:3562"
              /tissue_type="leaves"
              /clone_lib="lambda gt11"
              /clone="pALC 18"
              22..1095
              /EC_number="4.1.2.13"
              /codon_start=1
              /product="fructose-bisphosphate aldolase"
              /protein_id="CAA46649.1"
              /db_xref="GI:22620"
              /translation="WTAVRGKVADELIANASYIATPGKVILAADESTCTIGKRPPSIN
              VENVESNRRLRELFTTPGALPYLSGVILEETLYOKTADGKPFVDAMKGGVLPGI
              KYDGLVLAGINGETTTQGLDGLAQRCQYITAGARFAKRAVLKIGTPEPPLAIL
              ENAGLARYGICENGVLPIVEPEILVDGTHIDRCAEVSEVILAAICYKALNDHVL
              LEGTSLKNIVTPGSESKVPTEVIAETVTLQRTVPQAVPGVMFLSGGQSEEEATL
              NLNANKLETKPTWLTSPSYGRALQSTLKAWQGEENVAKAQEVLARAKGNSEATL
              GKYQGAGGADASESLHKDYKY"
              poly_site 1377
BASE COUNT   370 a 304 c 332 g 384 t
ORIGIN
Query Match      2.5%; Score 21; DB 7; Length 1390;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 833 ttctgcgaaaaaataaaaaa 853
Db 1366 TTTTGGCAAAAAAATAAAAA 1386

LOCUS       HB080598      1493 bp      mRNA           PLN      29-JUL-1998
DEFINITION   Hevea brasiliensis latex patatin homolog mRNA, complete cds.
ACCESSION    HB080598
VERSION      U80598.1 GI:1916804
KEYWORDS     Para rubber tree.
SOURCE       Hevea brasiliensis
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
              eudicots; Rosidae; eurosids I; Malpighiales; Euphorbiaceae; Hevea.
              1 (bases 1 to 1493)
              AUTHORS      Beezhoid,D.H., Sussman,G.L., Liss,G.M. and Chang,N.S.
              TITLE        Latex allergy can induce clinical reactions to specific foods
              JOURNAL      Clin. Exp. Allergy 26 (4), 416-422 (1996)
FEATURES     source
              2 (bases 1 to 1493)
              AUTHORS      Kostyal,D.A., Hickey,V.L., Noti,J.D., Sussman,G.L. and
              Beezhoid,D.H.
              TITLE        Cloning and characterization of a latex allergen (Hev b 7):
              JOURNAL      homology to patatin, a plant PLA2
              MEDLINE      Clin. Exp. Immunol. 112 (3), 355-362 (1998)
              AUTHORS      Kostyal,D.A., Beezhoid,D.H., Hickey,V.L., Noti,J.D. and
              Sussman,G.L.
              TITLE        Direct Submission
              JOURNAL      Submitted (29-NOV-1996) Macrophage Biology, Guthrie Research
              Institute, 1 Guthrie Square, Sayre, PA 18840, USA
              FEATURES     Location/Qualifiers
              1..1493
              /organism="Hevea brasiliensis"
              /db_xref="taxon:3981"
              69..1235
              /note="putative PLA2: latex protein allergen; similar to
              Solanum tubulin patatin encoded by GenBank Accession
              Number X03932"
              /codon_start=1
              /product="latex patatin homolog"
              /protein_id="AAC27724.1"
              /db_xref="GI:1916805"
              /translation="MATGSTLTQCKKITVLSIDGGIRGIPGIIASLESKLQDL
              GPDRIADYDFIIAGTSTGGLITMTAPNEDKKPMYQAKDKDFYLCNCKRIFPKES
              RDNDYPIHSIGVIGVEYLRELCNNLLKDLAVKDTSTDIPTFDIKLLLPVIFPSDD
              AKCNALKNARLADVCISTSAAPVLLPAHSFTEDDKNHTFELIDGGVAATNPTLLAL
              THIRNEIIRQNPRIGANLTESKSLVSLGTSKSEYKEKNADMTSKRWLYNWLALYN
              GNSPAVDIFSASDMVDHSLFSLKDCEDYLRIOODTLTGESGHATENLQ
              RLVEIGTELEKQESRLNLTGRLSEIFGAPTNEAAIAFAKLLSEERKLQLK"
BASE COUNT   492 a 289 c 301 g 411 t
ORIGIN
Query Match      2.5%; Score 21; DB 8; Length 1493;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 833 ttctgcgaaaaaataaaaaa 853
Db 1457 TTTTGGCAAAAAAATAAAAA 1477

LOCUS       HSM801097      2424 bp      mRNA           PRI      15-SEP-1999
DEFINITION   Homo sapiens mRNA; cDNA DKFZp564F0416 (from clone DKFZp564F0416).
ACCESSION    AL117570
VERSION      AL117570.1 GI:5912122
KEYWORDS     human.
SOURCE       Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 2424)
              AUTHORS      Blum,H., Bauersachs,S., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
              TITLE        Direct Submission
              JOURNAL      Submitted (15-SEP-1999) MIPS, Am Klopferspitz 18a D-82152
              Martinsried, GERMANY
              COMMENT      Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
              Research Center (DKFZ); sequenced by IMG within the cDNA sequencing
              consortium of the German Genome Project. This clone is available at
              the RZPD in Berlin.
              Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
              Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
              FEATURES     Location/Qualifiers
              1..2424
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="DKFZp564F0416"
BASE COUNT   630 a 304 c 332 g 384 t
ORIGIN
Query Match      2.5%; Score 21; DB 7; Length 1390;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 833 ttctgcgaaaaaataaaaaa 853
Db 1366 TTTTGGCAAAAAAATAAAAA 1386
```

polyA\_site 583 a 525 c 414 g 802 t  
 BASE COUNT 683 a 525 c 414 g 802 t  
 ORIGIN  
 /clone\_lib="564 (synonym: hfr2). Vector pAMP1; host  
 xl-2blue; sites NotI + SalI"  
 /dev\_stage="fetal"  
 /tissue\_type="brain"  
 2403

Query Match 2.5%; Score 21; DB 10; Length 2424;  
 Best Local Similarity 100.0%; Pred. No. 1.6;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 833 ttttcgcaaaaaa 853  
 |  
 Db 2397 TTTTCGCAAAAAA 2417

RESULT 10  
 CELH43107/c  
 LOCUS CELH43107 26894 bp DNA INV 11-FEB-1999  
 DEFINITION Caenorhabditis elegans cosmid H43107.  
 ACCESSION AF125965  
 VERSION AF125965.1 GI:4262641  
 KEYWORDS  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.

REFERENCE 1 (bases 1 to 26894)  
 AUTHORS Wilton, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightings, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A., Saunders, D., Showkneen, R., Smalton, N., Smith, A., Sonhammer, E., Vaughan, K., Waterston, R., Thierly-Mieg, J., Thomas, K., Vaudin, M., Wilkinson-Sprat, J., Watson, A., Weinstock, L., Wilkinon-Sprat, J. and Wohldman, P.  
 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans  
 Nature 368 (6466), 32-38 (1994)  
 94150718  
 2 (bases 1 to 26894)  
 Cordes, M. and Broy, M.  
 The sequence of C. elegans cosmid H43107  
 Unpublished (1999)  
 3 (bases 1 to 26894)  
 Waterston, R.  
 Direct Submission  
 Submitted (04-FEB-1999) Department of Genetics, Washington University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 Submitted by:  
 Genome Sequencing Center  
 Department of Genetics, Washington University,  
 St. Louis, MO 63110, USA, and  
 Sanger Centre, Hinxton Hall  
 Cambridge CB10 1RQ, England  
 e-mail: rw@nemato.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 3' clone is E03D2, 200 bp overlap. Actual start of this clone is at base position 1 of CELH43107; actual end is at 8142 of CELE03D2

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES

Location/Qualifiers  
 1. .26894  
 /organism="Caenorhabditis elegans"  
 /strain="Bristol N2"  
 /db\_xref="taxon:6239"  
 /clone="H43107"  
 /chromosome="v"  
 3701. .11644  
 /gene="H43107.2"  
 join(3701. .3795,3846. .4029,4096. .4288,5106. .5277,5327. .5480,6310. .6498,7570. .7659,8379. .9058,10102. .10205,10258. .10374,11529. .11644)  
 /gene="H43107.2"

gene

CDS

source

/note="similar to glycosyl transferases (Pfam:PF00535, Score:55.6, E= 1.1e-12, N=1); coded for by C. elegans cDNA yk28d11.3; coded for by C. elegans cDNA yk258a7.3; coded for by C. elegans cDNA yk258a7.5; coded for by C. elegans cDNA yk265h10.5; coded for by C. elegans cDNA yk187a1.5; coded for by C. elegans cDNA yk187a1.5"  
 /codon\_start=1  
 /protein\_id="AAD14758.1"  
 /db\_xref="GI:4262643"  
 /translation="WDLLSPFISQNLCLGLFLVASIVSSLTLSYATPWRRRR  
 AYLPEVIRSPQKPFVLLQEDRDAPKRTTPSCYLSVLIIPAMNEVEREIMDD  
 CDDYLEARAEKDKFTYIIIVDDGSTDETADIVQIGARRONLRVKKMNRGGGA  
 VKMGLVHSSGKILFADAGATKFAFENLEKEMLTAGGELDFSFVIVASRAHL  
 EASMAVRSVPRTILMGLFHLIVYFAARTIKDTQCFLFRTSAAIRFPVHLIERW  
 AFDVLIFLCRWTPVKEVSVRVTEIGSKITPWSLQMGDRDLVLWFRYAKMVK  
 KAAATSKOATVHSDIEKVEVLTDECTDYVEDDTKOFDINEYCDQIOAVMVE  
 SDGMTELFEDIRIAPIANALRRVLIAEVPTMALEKIYLYONTSVIQDEVLCRLGL  
 LPLRVDPGFQPKKEVGINEKGVDCDEEPPGPAKNLFIKINVSCKNNRNPATAT  
 DPQLYHNSVSFRAPFWPTADQKTFTEAHPRMVSDDLVAKLRFQETEAACHA  
 VKGIGRDHAKSPVATASYRLPTIRLNAEISGEAAERLKSVSFSGVIAIEKGAKRI  
 AVVKDARKDTCRNVRHEDLSKVVLQGNKQHFIFSVESGALKSELVVEACKVME  
 IKRSRLRQIEALIQ"  
 12875. .16451  
 /gene="H43107.1"  
 join(12875. .12980,13557. .13795,14576. .14828,16249. .16451)  
 /gene="H43107.1"  
 /note="coded for by C. elegans cDNA yk55h6.5"  
 /codon\_start=1  
 /protein\_id="AAD14757.1"  
 /db\_xref="GI:4262642"  
 /translation="MIAAQVSKPIADAVLRVYKGDHPVRNKLPIGRGLVRFVRSRL  
 RMRKLGEPISQAPISSEAALEQVQVQLVFSYSGVGFAGYFYTKMTTPESLK  
 IEDYEFQQRQAIKELRLQVLESLQAQKRNFFSQLGFKENGDESPKPSQQ  
 APPEKPSPEKRPATSELRKFORISLPIIDAAALILDDDEYLVVRGSSVGIELSV  
 NFAAPGDRILRVIPSPKLRSCAOMAKENACHVIGAKDRSFRVNS"

gene

CDS

BASE COUNT 8709 a 4846 c 4398 g 8941 t  
 ORIGIN

Query Match 2.5%; Score 21; DB 35; Length 26894;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 833 ttttcgcaaaaaa 853  
 |  
 Db 11053 TTTTCGCAAAAAA 11033



```

RESULT 11
AC017014_3/c
WPCOMMENT
Sequence split into 4 fragments LOCUS AC017014 Accession AC017014
Fragment Name Begin End
AC017014_0 1 110000
AC017014_1 100001 210000
AC017014_2 200001 310000
AC017014_3 300001 367220
Continuation (4 of 4) of AC017014 from base 300001 (AC017014 Homo sapiens clone RP11-142)

Query Match 2.5%; Score 21; DB 45; Length 67220;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 833 ttttcgaaaaaataaaaaa 853
|||||
DB 4455 TTTTCGAAAAAATAAAAAA 4435

RESULT 12
AC017014_2/c
WPCOMMENT
Sequence split into 4 fragments LOCUS AC017014 Accession AC017014
Fragment Name Begin End
AC017014_0 1 110000
AC017014_1 100001 210000
AC017014_2 200001 310000
AC017014_3 300001 367220
Continuation (3 of 4) of AC017014 from base 200001 (AC017014 Homo sapiens clone RP11-142)

Query Match 2.5%; Score 21; DB 45; Length 110000;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 833 ttttcgaaaaaataaaaaa 853
|||||
DB 104455 TTTTCGAAAAAATAAAAAA 104435

RESULT 13
AC008878/c
LOCUS AC008878 291288 bp DNA HTG 31-OCT-1999
DEFINITION Homo sapiens chromosome 19 clone C19B-H1_2207023, *** SEQUENCING IN
PROGRESS ***, 160 unordered pieces.
ACCESSION AC008878
VERSION AC008878.2 GI:6165135
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 291288)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 291288)
DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Oct 31, 1999 this sequence version replaced gi:5686195.
www.jgi.doe.gov.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 160 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

```

* be preserved.
* 1
* 1016: contig of 1016 bp in length
* gap of unknown length
* 1017: contig of 585 bp in length
* gap of unknown length
* 1602: contig of 826 bp in length
* gap of unknown length
* 2428: contig of 656 bp in length
* gap of unknown length
* 3084: contig of 876 bp in length
* gap of unknown length
* 3960: contig of 3921 bp in length
* gap of unknown length
* 7881: contig of 696 bp in length
* gap of unknown length
* 8577: contig of 3824 bp in length
* gap of unknown length
* 12401: contig of 810 bp in length
* gap of unknown length
* 13211: contig of 719 bp in length
* gap of unknown length
* 13930: contig of 2962 bp in length
* gap of unknown length
* 16892: contig of 631 bp in length
* gap of unknown length
* 17523: contig of 532 bp in length
* gap of unknown length
* 18055: contig of 682 bp in length
* gap of unknown length
* 18737: contig of 883 bp in length
* gap of unknown length
* 19620: contig of 847 bp in length
* gap of unknown length
* 20466: contig of 943 bp in length
* gap of unknown length
* 21409: contig of 891 bp in length
* gap of unknown length
* 22301: contig of 963 bp in length
* gap of unknown length
* 23264: contig of 1007 bp in length
* gap of unknown length
* 24271: contig of 744 bp in length
* gap of unknown length
* 25015: contig of 651 bp in length
* gap of unknown length
* 25666: contig of 1425 bp in length
* gap of unknown length
* 27091: contig of 975 bp in length
* gap of unknown length
* 28066: contig of 772 bp in length
* gap of unknown length
* 28838: contig of 3510 bp in length
* gap of unknown length
* 32348: contig of 771 bp in length
* gap of unknown length
* 33119: contig of 639 bp in length
* gap of unknown length
* 33758: contig of 864 bp in length
* gap of unknown length
* 34622: contig of 536 bp in length
* gap of unknown length
* 35158: contig of 1039 bp in length
* gap of unknown length
* 36197: contig of 1137 bp in length
* gap of unknown length
* 37334: contig of 721 bp in length
* gap of unknown length
* 38055: contig of 708 bp in length
* gap of unknown length
* 38763: contig of 762 bp in length
* gap of unknown length
* 39525: contig of 652 bp in length
* gap of unknown length

```

```
* 40177 42851: contig of 2675 bp in length
*      gap of unknown length
* 42852 43760: contig of 909 bp in length
*      gap of unknown length
* 43761 44291: contig of 531 bp in length
*      gap of unknown length
* 44292 44840: contig of 549 bp in length
*      gap of unknown length
* 44841 45627: contig of 787 bp in length
*      gap of unknown length
* 45628 46298: contig of 671 bp in length
*      gap of unknown length
* 46299 47167: contig of 869 bp in length
*      gap of unknown length
* 47168 47878: contig of 711 bp in length
*      gap of unknown length
* 47879 48711: contig of 833 bp in length
*      gap of unknown length
* 48712 48959: contig of 248 bp in length
*      gap of unknown length
* 48960 50309: contig of 1350 bp in length
*      gap of unknown length
* 50310 51028: contig of 719 bp in length
*      gap of unknown length
* 51029 51872: contig of 844 bp in length
*      gap of unknown length
* 51873 52523: contig of 651 bp in length
*      gap of unknown length
* 52524 53417: contig of 894 bp in length
*      gap of unknown length
* 53418 54178: contig of 761 bp in length
*      gap of unknown length
* 54179 54928: contig of 750 bp in length
*      gap of unknown length
* 54929 55629: contig of 701 bp in length
*      gap of unknown length
* 55630 56741: contig of 1112 bp in length
*      gap of unknown length
* 56742 58251: contig of 1510 bp in length
*      gap of unknown length
* 58252 59103: contig of 852 bp in length
*      gap of unknown length
* 59104 60607: contig of 1504 bp in length
*      gap of unknown length
* 60608 61642: contig of 1035 bp in length
*      gap of unknown length
* 61643 62432: contig of 790 bp in length
*      gap of unknown length
* 62433 62684: contig of 252 bp in length
*      gap of unknown length
* 62685 64028: contig of 1344 bp in length
*      gap of unknown length
* 64029 68403: contig of 4375 bp in length
*      gap of unknown length
* 68404 69905: contig of 1502 bp in length
*      gap of unknown length
* 69906 70653: contig of 748 bp in length
*      gap of unknown length
* 70654 72376: contig of 1723 bp in length
*      gap of unknown length
* 72377 73174: contig of 798 bp in length
*      gap of unknown length
* 73175 73783: contig of 609 bp in length
*      gap of unknown length
* 73784 74652: contig of 869 bp in length
*      gap of unknown length
* 74653 75356: contig of 704 bp in length
*      gap of unknown length
* 75357 76056: contig of 700 bp in length
*      gap of unknown length
* 76057 76965: contig of 909 bp in length
*      gap of unknown length
* 76966 77981: contig of 1016 bp in length
```

```
*      gap of unknown length
* 77982 79239: contig of 1258 bp in length
*      gap of unknown length
* 79240 80222: contig of 983 bp in length
*      gap of unknown length
* 80223 81434: contig of 1212 bp in length
*      gap of unknown length
* 81435 82163: contig of 729 bp in length
*      gap of unknown length
* 82164 82879: contig of 716 bp in length
*      gap of unknown length
* 82880 83520: contig of 641 bp in length
*      gap of unknown length
* 83521 83612: contig of 92 bp in length
*      gap of unknown length
* 83613 83759: contig of 147 bp in length
*      gap of unknown length
* 83760 84424: contig of 665 bp in length
*      gap of unknown length
* 84425 85182: contig of 758 bp in length
*      gap of unknown length
* 85183 86189: contig of 1007 bp in length
*      gap of unknown length
* 86190 87102: contig of 913 bp in length
*      gap of unknown length
* 87103 87755: contig of 653 bp in length

Query Match      2.5%; Score 21; DB 42; Length 291288;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 833 ttctgcgaataaaataaa 853
      |||||
Db 212340 TTTTGGCAAAAAAAAAA 212320
```

## RESULT 14

```
AF047051 LOCUS 297 bp mRNA PLN 21-FEB-1998
Glycine max ribosomal protein L41 mRNA, partial cds.
```

```
AF047051 DEFINITION
AF047051 ACCESSION
AF047051.1 GI:2905777
```

## KEYWORDS

soybean.

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;

Glycine.

1 (bases 1 to 297)

Woo.H.-H.

Primary structure of mRNA encoding a ribosomal protein L41

Unpublished

2 (bases 1 to 297)

Mahalingam,R. and Knap,H.T.

Direct Submission

Submitted (09-FEB-1998) Agronomy, Clemson University, P & A Building, Clemson, SC 29634, USA

## FEATURES

source

1..297

/organism="Glycine max"

/cultivar="Essex"

/db\_xref="taxon:3847"

/tissue\_type="roots"

/clone="CP-2"

/note="PI 437654"

<1..78

/codon\_start=1

/product="ribosomal protein L41"

/protein\_id="AAC03557.1"

/db\_xref="GI:2905778"

/translation="MGVGKTKRMRLKRRKMRQRSK"

CDS

86 a 41 c 73 g 97 t

BASE COUNT

ORIGIN

Query Match 2.3%; Score 20; DB 8; Length 297;  
Best Local Similarity 100.0%; Pred. No. 6.9;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 834 ttgcgaaaaaaaaaaaaa 853  
|||||  
DB 275 TTTGCGAAAAAAAAAAAA 294

RESULT 15

CEU82968 949 bp mRNA INV 15-JAN-1997  
LOCUS Caenorhabditis elegans MAD-like-1 homolog (mdl-1) mRNA, complete  
DEFINITION cds.  
ACCESSION U82968  
VERSION U82968.1 GI:1778833  
KEYWORDS  
SOURCE  
ORGANISM Caenorhabditis elegans.  
Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 949)  
Yuan, J., Tirabassi, R., Bush, A. and Cole, M.D.  
The C. elegans homologues of the vertebrate  
proliferation/differentiation-associated genes MAD/MXII and MAX are  
expressed in a dynamic pattern in development  
Unpublished  
2 (bases 1 to 949)  
Yuan, J. and Cole, M.D.  
Direct Submission  
TITLE Submitted (20-DEC-1996) Molecular Biology, Princeton University,  
Lewis Thomas Laboratory, Princeton, NJ 08544, USA  
JOURNAL Location/Qualifiers  
REFERENCE 1..949  
/organism="Caenorhabditis elegans"  
/strain="N2"  
/db\_xref="taxon:6239"  
/chromosome="X"  
/map="between unc-6 and sup-7"  
/note="sequence also located within cosmid R03E9, GenBank  
Accession Number U40947"  
1..846  
/gene="mdl-1"  
1..846  
/gene="mdl-1"  
/function="dimerizes with MXL-1 to bind DNA"  
/note="MDL1; basic helix-loop-helix leucine zipper DNA  
binding protein; similar to mouse Max interacting protein  
1: Swiss-Prot Accession Number P50540"  
/codon\_start=1  
/product="MAD-like-1 homolog"  
/protein\_id="AA040927.1"  
/db\_xref="GI:1778834"

FEATURES

source  
1..949  
/organism="Caenorhabditis elegans"  
/strain="N2"  
/db\_xref="taxon:6239"  
/chromosome="X"  
/map="between unc-6 and sup-7"  
/note="sequence also located within cosmid R03E9, GenBank  
Accession Number U40947"  
1..846  
/gene="mdl-1"  
1..846  
/gene="mdl-1"  
/function="dimerizes with MXL-1 to bind DNA"  
/note="MDL1; basic helix-loop-helix leucine zipper DNA  
binding protein; similar to mouse Max interacting protein  
1: Swiss-Prot Accession Number P50540"  
/codon\_start=1  
/product="MAD-like-1 homolog"  
/protein\_id="AA040927.1"  
/db\_xref="GI:1778834"  
/translation="MEQQLNGLHTAARLLDIGALDGLSSLDLGLTSSSPGSSSP  
AMFDLSNSELRLSCGKLVKQKSCASNASTSQPYCSPPPARKSKHSRTAHNE  
LEKTRANLRCLETKMLVPCVSDATNTVIALTRARDHIEIQDSNAQMKKIND  
LRDEQELVAELQAEVAQAATSQACQTLQSQRSPESRASFTSTSSRDSPCYLEY  
SPSSKPMDSHKPTIIDLYAEGLI PRGPIFFPRPLVYPHNVFDLMLPPTPDVSQFLP  
INLQV"

BASE COUNT 279 a 277 c 156 g 237 t  
ORIGIN

Query Match 2.3%; Score 20; DB 34; Length 949;  
Best Local Similarity 100.0%; Pred. No. 6.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 834 ttgcgaaaaaaaaaaaaa 853  
|||||  
DB 923 TTTGCGAAAAAAAAAAAA 942

Search completed: May 1, 2000, 16:53:11  
Job time: 21291 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 18:53:57 ; Search time 250.34 Seconds  
(without alignments)  
852.496 Million cell updates/sec

```

Title: US-09-215-435-123
Perfect score: 853
Sequence: 1 ggagatggcgagcagtct.....tttgcgaaaaaaaaaaaaa 853

```

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 311585 seqs, 125096042 residues

```

Word size :      0
Total number of hits satisfying chosen parameters: 623170

```

```
Minimum DB seq length: 0
Maximum DB seq length: 1000000
```

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %			DB	ID	Description
		Match	Length				
1	170	19.9	296	1	V89317	EST clone C1437. N	
2	138	15.0	239	1	X40653	Human secreted pro	
3	90	10.6	184	1	X39568	Human secreted pro	
C 4	21	2.5	27	1	X23566	Deletion sequence	
5	21	2.5	910	1	V66759	Pathogen response	
C 6	20	2.3	1419	1	T10928	Nucleotide sequenc	
C 7	20	2.3	1419	1	V62709	Maize male fertili	
C 8	20	2.3	1419	1	V73926	Maize MS45 cDNA. N	
C 9	20	2.3	1419	1	X00482	Zea mays male fert	
10	20	2.3	4434	1	T93389	Bloom's syndrome B	
11	20	2.3	4437	1	T67013	Bloom syndrome act	
12	20	2.3	4437	1	T93390	Bloom's syndrome B	
13	20	2.3	4437	1	T93392	Bloom's syndrome B	
14	20	2.3	4437	1	T93394	Bloom's syndrome B	
15	20	2.3	4437	1	T93395	Bloom's syndrome B	
16	20	2.3	4438	1	T93391	Bloom's syndrome B	
17	20	2.3	4438	1	T93393	Bloom's syndrome B	
18	19	2.2	1031	1	Q99786	Plant SAR gene pDP	
19	19	2.2	1031	1	V81686	Arabidopsis protei	
20	19	2.2	1031	1	V62802	Tobacco SAR CHX in	
21	19	2.2	1288	1	V04075	Human cytokine/pep	
22	19	2.2	1288	1	V02295	Homo sapiens cDNA	
23	19	2.2	1288	1	V04131	Human HR-1 recepto	
C 24	19	2.2	1336	1	V59727	Human secreted pro	
25	19	2.2	1369	1	T95214	cDNA encoding the	
26	19	2.2	1541	1	V59565	Human secreted pro	
27	19	2.2	2160	1	T85328	Truncated murine T	
C 28	19	2.2	4796	1	T76897	Cauliflower floral	
C 29	19	2.2	4816	1	V58305	Brassica oleracea	
C 30	18	2.1	27	1	Q75748	Reverse transcript	
C 31	18	2.1	21	1	X23565	Deletion sequence	
32	18	2.1	161	1	V00416	3' fragment of clo	
33	18	2.1	1136	1	X25660	Human endogenous r	

## ALIGNMENTS

```

RESULT 1
V89317 standard; cDNA; 296 BP.
ID V89317;
AC 15-FEB-1999 (first entry)
DE EST clone C1437
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW gene therapy; ss.
KW Homo sapiens.
PN WO9845436-A2.
PS 15-OCT-1998.
PD 10-APR-1998; U06955.
PR 10-APR-1997; US-838821.
PR (GEMX) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
PI WPI: 99-070077/06.
DR New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
PS Claim 1; Page 180; 619pp; English.
CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
CC Sequence 296 BP; 56 A; 70 C; 72 G; 98 T;
SQ
Query Match 19.9%; Score 170; DB 1; Length 296;
Best Local Similarity 100.0%; Pred. No. 2.5e-66;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 402 ccacaatagcgggattgatctccttagacctatcttggcgttgccaggtccctttacc 461
DB 2 CCACAAATAGCGGGATTGATCTCTTTAGACCTATCTTTGGCGTTGCCAGTTCCTTTACC 61
QY 462 tttgtgagtttagttgattgcttcttgggcgtttgatcgaccttgccttgcattg 521
DB 62 TTTTGTGAGTTAGTTTGTGCTCTTTGGGCTTTGATCGGACTTTGTGCTTGCAATTG 121
QY 522 ccgaagcttatatccaccattgcccagcggtattctccatctcctttgcag 571
DB 122 CCGAAGCTTATATCCACCAATTGCCACGGGCAATTCCTCATCTCCTTGCAG 171
RESULT 2

```



CC sample of target deletion oligonucleotides which are labelled and  
 CC hybridize with the probe oligonucleotides of the sensor arrays. Such  
 CC oligonucleotides and their targets are represented in X23548-X23709.  
 CC Oligonucleotides characterized by the method form pharmaceutical  
 CC compositions that are useful for modulating cellular adhesion or  
 CC proliferation, and being active against a eukaryotic pathogen, a human  
 CC retrovirus, a human immunodeficiency virus (HIV), or a non-human  
 CC retrovirus, including influenza virus, Epstein-Barr virus, Respiratory  
 CC Syncytial virus or cytomegalovirus (CMV). The compositions enable  
 CC characterization of deletion sequence oligonucleotides having related,  
 CC but different nucleobase sequences, and quantification of different  
 CC species of deletion sequence ("target") oligonucleotides in a mixture.  
 CC Also, if the specificity of the oligonucleotide's nucleobase sequence  
 CC for its reverse complement is not modified, the method may be performed  
 CC using oligodeoxynucleotides.  
 CC Sequence 27 BP; 7 A; 2 C; 3 G; 15 T;

Query Match 2.5%; Score 21; DB 1; Length 27;  
 Best Local Similarity 100.0%; Pred. No. 2.2; Mismatches 0; Indels 0; Gaps 0;  
 Matches 21; Conservative 0;

QY 833 ttctgcgaaaaa 853  
 Db 22 TTTTGGAAAAA 2

RESULT 5  
 V66759  
 ID V66759 standard; DNA; 910 BP.  
 AC V66759;  
 DT 02-FEB-1999 (first entry)  
 DE Pathogen response protein LSD1-interacting protein HH DNA.  
 KW LSD1-interacting protein HH; plant pathogen response; apoptosis;  
 KW programmed cell death; disease resistance; herbicide resistance;  
 KW transgenic plant; crop protection; ss.  
 OS Arabidopsis thaliana  
 FH Key Location/Qualifiers  
 FT CDS 1..696  
 FT /\*tag= a

PN W09837755-A1.  
 PD 03-SEP-1998.  
 PE 27-FEB-1998: U04077.  
 PR 28-FEB-1997: US-039063.  
 PA (OYMC-) UNIV NORTH CAROLINA.  
 PI Dangl JL, Dietrich RA, Epple PM, Richberg MH;  
 DR WPI: 98-531501/45.  
 DR P-PSDB: W72388.  
 PT New isolated Arabidopsis genes - useful for producing transgenic  
 PT plants which show resistance to cell death caused by pathogens or  
 PT herbicides.  
 PS Claim 46; Page 57-58; 89pp; English.  
 CC This is the nucleotide sequence of DNA encoding LSD1-interacting  
 CC protein HH (see W72388) of Arabidopsis thaliana. LSD1 interacting  
 CC genes (see V66755-67) were isolated from a yeast gene expression  
 CC library constructed in plasmid pJG4-5 using RNA from Arabidopsis  
 CC leaves infected with Pseudomonas syringae. A two-hybrid system was  
 CC used with LSD1 short and long open reading frames (see V66750-51)  
 CC as bait. LSD1 (see W72366-67) is a novel polypeptide that regulates  
 CC the initial response of plants to pathogens and the subsequent  
 CC spread of plant cell death engendered by infection. Because the  
 CC inactivation of LSD1 by mutation leads to enhanced disease  
 CC resistance, LSD1 partner proteins represent novel targets for  
 CC engineering plants with enhanced resistance to pathogens. Thus,  
 CC the invention includes all proteins (see W72384-96) that interact  
 CC with the cell death regulator LSD1.  
 CC Sequence 910 BP; 265 A; 159 C; 238 G; 248 T;

Query Match 2.5%; Score 21; DB 1; Length 910;  
 Best Local Similarity 100.0%; Pred. No. 1.9; Mismatches 0; Indels 0; Gaps 0;  
 Matches 21; Conservative 0;

QY 833 ttctgcgaaaaa 853  
 Db 886 TTTTGGAAAAA 906

RESULT 6  
 T10928/c  
 ID T10928 standard; DNA; 1419 BP.  
 AC T10928;  
 DT 10-JUL-1996 (first entry)  
 DE Nucleotide sequence for mediating male fertility in plants.  
 KW Male fertility; plant; microsporogenesis; tassel development; MS45;  
 KW maize; sterile; fertile; transformed plant; female parent;  
 KW hybrid seed; ds.  
 OS Zea mays.  
 PN US5478369-A.  
 PD 26-DEC-1995.  
 PE 12-JUN-1990; 537183.  
 PR 12-JUN-1990; US-537183.  
 PR 02-AUG-1993; US-103739.  
 PR 28-OCT-1994; WO-012444.  
 PA (PION-) PIONEER HI-BRED INT INC.  
 PI Albertsen MC, Beach LR, Howard J, Huffman GA;  
 DR WPI: 96-057646/06.  
 DR P-PSDB: R88502.  
 DT Nucleic acid encoding a protein critical for male fertility in  
 PT plants - used to produce plants, esp. maize, that are normally male  
 PT sterile but can be induced to fertility, esp. for use in hybrid seed  
 PT prodn

PS Claim 4; Column 25-27; 27pp; English.  
 CC This sequence represents a nucleic acid sequence which mediates male  
 CC fertility in plants. This sequence encodes an amino acid sequence  
 CC which is responsible for one of the steps in microsporogenesis,  
 CC specifically tassel development. This cDNA was isolated from a  
 CC tassel derived cDNA and was named MS45. The introduction of this  
 CC cDNA into a plant, pref. maize, which is normally male sterile  
 CC causes it to be fertile. Such transformed plants may be used as  
 CC female parents in the production of hybrid seeds.  
 CC Sequence 1419 BP; 304 A; 387 C; 473 G; 255 T;

Query Match 2.3%; Score 20; DB 1; Length 1419;  
 Best Local Similarity 100.0%; Pred. No. 5; Mismatches 0; Indels 0; Gaps 0;  
 Matches 20; Conservative 0;

QY 105 ctccattggcacagacttct 124  
 Db 430 CTCATTGGCACAGACTTCT 411

RESULT 7  
 V62709/c  
 ID V62709 standard; cDNA; 1419 BP.  
 AC V62709;  
 DT 24-DEC-1998 (first entry)  
 DE Maize male fertility gene MS45.  
 KW ds; maize; male fertility gene; MS45; transgenic plant; hybrid seed.  
 OS Zea mays.  
 FH Key Location/Qualifiers  
 FT CDS 1..1419  
 FT /\*tag= a  
 FT /product= "MS45"  
 FT /note= "No start or stop codon given"  
 PN US5824524-A.  
 PD 20-OCT-1998.  
 PE 07-JUN-1995; 474404.  
 PR 02-AUG-1993; US-103739.  
 PR 12-JUN-1990; US-537183.  
 PR 07-JUN-1995; US-474404.  
 PA (PION-) PIONEER HI-BRED INT INC.  
 PI Albertsen MC, Beach LR, Howard J, Huffman GA;  
 DR WPI: 98-582558/49.  
 DR P-PSDB: W77413.

PT Production of male-sterile plants - by repressing expression of male  
 PT fertility gene  
 PS Claim 2; Column 37-40; 40pp; English.  
 CC The maize male fertility gene MS45 can be repressed as a method for  
 CC mediating male fertility in a plant. This method can be used especially  
 CC for producing hybrid maize seed. The endogenous gene can be inactivated  
 CC and the control of an inducible promoter can be used to specifically  
 CC activate the gene when fertile plants are desired.  
 SQ Sequence 1419 BP; 304 A; 387 C; 473 G; 255 T;

Query Match 2.3%; Score 20; DB 1; Length 1419;  
 Best Local Similarity 100.0%; Pred. No. 5;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 ctccattggcacagacttct 124  
 |||||  
 Db 430 CTCATTGGCACAGACTTCT 411

RESULT 8  
 V73926/c  
 ID V73926 standard; cDNA; 1419 BP.  
 AC V73926;  
 DT 04-MAR-1999 (first entry)  
 DE Maize MS45 cDNA.  
 KW Corn; MS45; plant; fertility; gene inactivation; inducible promoter; ds.  
 OS Zea mays.  
 PN US850014-A.  
 PD 15-DEC-1998.  
 PF 07-JUN-1995; 485845.  
 PR 02-AUG-1993; US-103739.  
 PR 12-JUN-1990; US-537183.  
 PR 07-JUN-1995; US-485845.  
 PA (PION-) PIONEER HI-BRED INT INC.  
 PI Albertsen MC, Beach LR, Howard J, Huffman GA;  
 DR WPI: 99-094416/08.  
 DR P-PSDB; W90062.  
 DT Non-maize plant containing defined cDNA sequence - and plant  
 PT containing defined amino acid sequence  
 PS Claim 1; Column 31-34; 35pp; English.  
 CC This sequence encodes the maize MS45 protein. This protein is used in a  
 CC method in which the fertility of a plant is controlled by inactivating a  
 CC gene critical to fertility and inserting into the plant the critical  
 CC gene linked to an inducible promoter.  
 SQ Sequence 1419 BP; 304 A; 387 C; 473 G; 255 T;

Query Match 2.3%; Score 20; DB 1; Length 1419;  
 Best Local Similarity 100.0%; Pred. No. 5;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 ctccattggcacagacttct 124  
 |||||  
 Db 430 CTCATTGGCACAGACTTCT 411

RESULT 9  
 X00482/c  
 ID X00482 standard; cDNA; 1419 BP.  
 AC X00482;  
 DT 30-MAR-1999 (first entry)  
 DE Zea mays male fertility MS45 cDNA.  
 KW Zea mays; maize; male fertility; MS45; sterile plant; hybrid strain;  
 KW breeding; ds.  
 OS Zea mays.  
 FH Key  
 FT CDS  
 FT Location/Qualifiers  
 FT 1..1419  
 FT /\*tag= a  
 FT /transl\_except= (pos:1264..1266,aa:Xaa)  
 FT /transl\_except= (pos:1270..1272,aa:Xaa)  
 FT /transl\_except= (pos:1312..1314,aa:Xaa)  
 FT /note= "Xaa is a stop codon; no stop codon given at

the end of the sequence"

FT US5859341-A.  
 PN 12-JAN-1999.  
 PD 07-JUN-1995; 482714.  
 PR 02-AUG-1993; US-103739.  
 PR 12-JUN-1990; US-537183.  
 PR 07-JUN-1995; US-482714.  
 PA (PION-) PIONEER HI-BRED INT INC.  
 PI Albertsen MC, Beach LR, Howard J, Huffman GA;  
 DR WPI: 99-120032/10.  
 DR P-PSDB; W30612.  
 DT Constitutively male sterile plants - with inducible male fertility,  
 PT useful in hybrid breeding  
 PS Example 1; Column 31-34; 36pp; English.  
 CC A method has been developed of providing heritable, externally  
 CC controlled male fertility in plants. The method comprises: (a) cloning a  
 CC gene (I) that encodes a product (II) essential for microsporogenesis;  
 CC (b) linking (I) to an expression control sequence that includes an  
 CC inducible promoter, responsive to external controls; (c) rendering the  
 CC inoperative the native gene that encodes (II); and (d) inserting the  
 CC plant that is constitutively sterile but controllably male fertile. Also  
 CC described in the present invention are controllably male fertile plants  
 CC produced by the above method, their parts, cells, and seeds (and any  
 CC plants grown from these seeds, their parts and cells), and hybrid seeds  
 CC produced using these plants. The method produces plants that are useful  
 CC in breeding hybrid strains. The rendering of constitutively sterile  
 CC plants fertile can tolerate 70-80% failure of induction without a  
 CC significant reduction in seed yield (considerably higher than known  
 CC methods based on rendering constitutively fertile plants sterile), no  
 CC manual removal of tassels (from maize plants) is required, and no  
 CC treatment with chemicals is needed during hybrid development. The  
 CC present sequence represents MS45 cDNA isolated from Zea mays in an  
 CC example from the present invention.  
 SQ Sequence 1419 BP; 304 A; 387 C; 473 G; 255 T;

Query Match 2.3%; Score 20; DB 1; Length 1419;  
 Best Local Similarity 100.0%; Pred. No. 5;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 ctccattggcacagacttct 124  
 |||||  
 Db 430 CTCATTGGCACAGACTTCT 411

RESULT 10  
 T93389  
 ID T93389 standard; cDNA; 4434 BP.  
 AC T93389;  
 DT 27-JAN-1998 (first entry)  
 DE Bloom's syndrome BLM mutated gene (3 base deletion at position 631).  
 KW BLM; Bloom's syndrome; BS; mutant; mutation; deletion; truncation;  
 KW therapy; diagnosis; ss.  
 OS Homo sapiens.  
 FH Key  
 FT CDS  
 FT 75..633  
 FT /\*tag= a  
 FT /note= "encodes a truncated protein"  
 FT mutation  
 FT 631..632  
 FT /\*tag= b  
 FT /note= "Deletion of three bases (CAA) from the wild  
 FT type HI-5' sequence between these positions  
 FT results in a stop codon"

WO9717979-A1.  
 PN 22-MAY-1997.  
 PD 15-NOV-1996; U19046.  
 PR 15-NOV-1995; US-559303.  
 PA (NYBL-) NEW YORK BLOOD CENT INC.  
 PI Ellis N, German J, Groden J;  
 DR WPI: 97-289051/26.  
 DR P-PSDB; W31545.  
 PT Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM



PT genes - for gene therapy with nucleic acid encoding active BLM  
 PT protein to treat Bloom's syndrome and cancer in general  
 PS Disclosure; Page -; 51pp; English.  
 CC This is a mutated BLM gene sequence isolated from a Japanese  
 CC Bloom's Syndrome sufferer designated "97(Asok)". The deletion of  
 CC CAA from positions 631-633 of the wild-type gene (i.e. between  
 CC nucleotides 631 and 632 of the present sequence) results in a stop  
 CC codon at amino acid position 186.  
 CC This is one of the seven unique mutations which were identified in  
 CC a study of 10 people with Bloom's Syndrome. Based on the various  
 CC mutations, diagnostic tests for Bloom's Syndrome have been developed  
 CC which use standard sequence analysis techniques to detect the presence  
 CC of 2 mutated BLM genes or the absence of a wild-type BLM gene.  
 CC Note: The present sequence does not appear in the specification; it  
 CC has been made by modifying the H1-5' wild type BLM sequence which is  
 CC provided in Figure 2 (T67013).  
 SQ Sequence 4434 BP; 1463 A; 866 C; 911 G; 1194 T;

Query Match 2.3%; Score 20; DB 1; Length 4434;  
 Best Local Similarity 100.0%; Pred. No. 4.8;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 tgaagaagacttataatgatg 229  
 |||||  
 Db 3073 TGAAGAAGACTTATAATGATG 3092

RESULT 11  
 T67013  
 ID T67013 standard; cDNA; 4437 BP.  
 AC T67013;  
 DT 26-JAN-1998 (first entry)  
 DE Bloom syndrome active BLM gene.  
 KW BLM; Bloom syndrome; BS; mutant; probe; PCR primer; cancer;  
 KW therapy; diagnosis; vector; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 75..4328  
 FT /\*tag= a  
 FT /product= BLM\_protein  
 FT 4329..4416  
 FT /\*tag= b  
 FT /polyA\_signal 4391..4396  
 FT /\*tag= c  
 FT 4417  
 FT /\*tag= d  
 PN WO9717979-A1.  
 PD 22-MAY-1997.  
 PF 15-NOV-1996; U19046.  
 PR 15-NOV-1995; US-559303.  
 PA (NYBL-) NEW YORK BLOOD CENT INC.  
 PI Ellis N, German J, Groden J;  
 DR WPI: 97-289051/26.  
 DR P-PSDB; W15264.  
 PT Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM  
 PT genes - for gene therapy with nucleic acid encoding active BLM  
 PT protein to treat Bloom's syndrome and cancer in general  
 PS Claim 59; Fig 2; 51pp; English.  
 CC This BLM gene H1-5' sequence encodes an enzymatically active BLM protein.  
 CC or the absence of a wild-type BLM gene in the nucleic acid of a subject  
 CC is detected. The BLM gene is in the 250 kb region between two markers  
 CC binding a 1.3 cm region on chromosome 15. cDNA from this region was  
 CC isolated, a 847 bp fragment selected and used to screen a HeLa cDNA  
 CC library. The longest clone H1 was isolated and extended by PCR to isolate  
 CC this BLM gene sequence. Delivery of a functional BLM gene to bone marrow  
 CC cells, e.g. via the viral vectors, is used to treat or prevent the onset  
 CC of Bloom's syndrome. Identification of the BLM gene and its products  
 CC should assist in the development of therapeutic and diagnostic agents for  
 CC cancer.  
 SQ Sequence 4437 BP; 1465 A; 867 C; 911 G; 1194 T;

Query Match 2.3%; Score 20; DB 1; Length 4437;  
 Best Local Similarity 100.0%; Pred. No. 4.8;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 tgaagaagacttataatgatg 229  
 |||||  
 Db 3076 TGAAGAAGACTTATAATGATG 3095

RESULT 12  
 T93390  
 ID T93390 standard; cDNA; 4437 BP.  
 AC T93390;  
 DT 27-JAN-1998 (first entry)  
 DE Bloom's syndrome BLM mutated gene (A-T substitution at position 888).  
 KW BLM; Bloom's syndrome; BS; mutant; mutation; truncation;  
 KW substitution; therapy; diagnosis; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 75..890  
 FT /\*tag= a  
 FT /note= "encodes a truncated protein"  
 FT 888  
 FT /\*tag= b  
 FT /note= "substitution of A in the wild type  
 FT H1-5' sequence to T at this  
 FT position results in a stop codon"  
 PN WO9717979-A1.  
 PD 22-MAY-1997.  
 PF 15-NOV-1996; U19046.  
 PR 15-NOV-1995; US-559303.  
 PA (NYBL-) NEW YORK BLOOD CENT INC.  
 PI Ellis N, German J, Groden J;  
 DR WPI: 97-289051/26.  
 DR P-PSDB; W31546.  
 PT Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM  
 PT genes - for gene therapy with nucleic acid encoding active BLM  
 PT protein to treat Bloom's syndrome and cancer in general  
 PS Disclosure; Page -; 51pp; English.  
 CC This is a mutated BLM gene sequence isolated from a German Bloom's  
 CC Syndrome sufferer designated "112(Nasch)". The substitution of  
 CC A at the position 888 of the wild type H1-5' gene to T results  
 CC in a stop codon at amino acid position 272.  
 CC This is one of the seven unique mutations which were identified in  
 CC a study of 10 people with Bloom's Syndrome. Based on the various  
 CC mutations, diagnostic tests for Bloom's Syndrome have been developed  
 CC which use standard sequence analysis techniques to detect the presence  
 CC of 2 mutated BLM genes or the absence of a wild-type BLM gene.  
 CC Note: The present sequence does not appear in the specification; it  
 CC has been made by modifying the H1-5' wild type BLM sequence which is  
 CC provided in Figure 2 (T67013).  
 SQ Sequence 4437 BP; 1464 A; 867 C; 911 G; 1195 T;

Query Match 2.3%; Score 20; DB 1; Length 4437;  
 Best Local Similarity 100.0%; Pred. No. 4.8;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 tgaagaagacttataatgatg 229  
 |||||  
 Db 3076 TGAAGAAGACTTATAATGATG 3095

RESULT 13  
 T93392  
 ID T93392 standard; cDNA; 4437 BP.  
 AC T93392;  
 DT 27-JAN-1998 (first entry)  
 DE Bloom's syndrome BLM mutated gene (A-G substitution at position 2089).  
 KW BLM; Bloom's syndrome; BS; mutant; mutation; substitution;  
 KW therapy; diagnosis; ss.  
 OS Homo sapiens.

```

FH Key          Location/Qualifiers
FT CDS          75..4328
FT              /*tag= a
FT              /note= "encodes a mutated BLM protein."
FT mutation     2089
FT              /*tag= b
FT              /note= "Substitution of the bp A at this position in
FT              the wild type H1-5' BLM gene to the bp G results
FT              in the amino acid Arg at the position 672"
FT
PN WO9717979-A1.
PD 22-MAY-1997.
PF 15-NOV-1996; U19046.
PR 15-NOV-1995; US-559303.
PA (NYBL-) NEW YORK BLOOD CENT INC.
PI Ellis N, German J, Groden J;
DR WPI: 97-289051/26.
DR P-PSDB; W31548.
PT Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM
PT genes - for gene therapy with nucleic acid encoding active BLM
PT protein to treat Bloom's syndrome and cancer in general
PT Disclosure; Page -: 51pp; English.
CC This is a mutated BLM gene sequence isolated from an American/European
CC Bloom's Syndrome sufferer designated "139(Vikre)". The substitution of
CC the bp A at the position 2089 of the wild type H1-5' gene to the bp G
CC results in the amino acid Arginine at the amino acid position 672.
CC This is one of the seven unique mutations which were identified in
CC a study of 10 people with Bloom's Syndrome. Based on the various
CC mutations, diagnostic tests for Bloom's Syndrome have been developed
CC which use standard sequence analysis techniques to detect the presence
CC of 2 mutated BLM genes or the absence of a wild-type BLM gene.
CC Note: The present sequence does not appear in the specification; it
CC has been made by modifying the H1-5' wild type BLM sequence which is
CC provided in Figure 2 (T67013).
SQ Sequence 4437 BP; 1464 A; 867 C; 912 G; 1194 T;

Query Match 2.3%; Score 20; DB 1; Length 4437;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 tgaagaagacttataatgatg 229
   |||||
Db 3076 TGAAGAAGACTTATAATGATG 3095

RESULT 14
T93394
ID T93394 standard; cDNA; 4437 BP.
AC T93394;
DT 27-JAN-1998 (first entry)
DE Bloom's syndrome BLM mutated gene (T-C substitution at position 2596).
KW BLM; Bloom's syndrome; BS; mutant; mutation; substitution;
KW therapy; diagnosis; ss.
OS Homo sapiens.
FH Key          Location/Qualifiers
FT CDS          75..4328
FT              /*tag= a
FT              /note= "encodes a mutated BLM protein"
FT mutation     2596
FT              /*tag= b
FT              /note= "Substitution of the bp T at this position in
FT              the wild type H1-5' BLM gene to the bp C results
FT              in the amino acid Thr at the position 841"
FT
PN WO9717979-A1.
PD 22-MAY-1997.
PF 15-NOV-1996; U19046.
PR 15-NOV-1995; US-559303.
PA (NYBL-) NEW YORK BLOOD CENT INC.
PI Ellis N, German J, Groden J;
DR WPI: 97-289051/26.
DR P-PSDB; W31550.
PT Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM
PT genes - for gene therapy with nucleic acid encoding active BLM
PT protein to treat Bloom's syndrome and cancer in general
PT Disclosure; Page -: 51pp; English.
CC This is a mutated BLM gene sequence isolated from an Italian
CC Bloom's Syndrome sufferer designated "113(Dadem)". The substitution of
CC the bp G at the position 3238 of the wild type H1-5' gene to the bp C
CC results in the amino acid Serine at the amino acid position 1055.
CC This is one of the seven unique mutations which were identified in
CC a study of 10 people with Bloom's Syndrome. Based on the various
CC mutations, diagnostic tests for Bloom's Syndrome have been developed
CC which use standard sequence analysis techniques to detect the presence
CC of 2 mutated BLM genes or the absence of a wild-type BLM gene.
CC Note: The present sequence does not appear in the specification; it
CC has been made by modifying the H1-5' wild type BLM sequence which is
CC provided in Figure 2 (T67013).
SQ Sequence 4437 BP; 1465 A; 868 C; 910 G; 1194 T;

Query Match 2.3%; Score 20; DB 1; Length 4437;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 tgaagaagacttataatgatg 229
   |||||
Db 3076 TGAAGAAGACTTATAATGATG 3095

RESULT 15
T93395
ID T93395 standard; cDNA; 4437 BP.
AC T93395;
DT 27-JAN-1998 (first entry)
DE Bloom's syndrome BLM mutated gene (G-C substitution at position 3238).
KW BLM; Bloom's syndrome; BS; mutant; mutation; substitution;
KW therapy; diagnosis; ss.
OS Homo sapiens.
FH Key          Location/Qualifiers
FT CDS          75..4328
FT              /*tag= a
FT              /note= "encodes a mutated BLM protein"
FT mutation     3238
FT              /*tag= b
FT              /note= "Substitution of the bp G at this position in
FT              the wild type H1-5' BLM gene to the bp C results
FT              in the amino acid Ser at the position 1055"
FT
PN WO9717979-A1.
PD 22-MAY-1997.
PF 15-NOV-1996; U19046.
PR 15-NOV-1995; US-559303.
PA (NYBL-) NEW YORK BLOOD CENT INC.
PI Ellis N, German J, Groden J;
DR WPI: 97-289051/26.
DR P-PSDB; W31551.
PT Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM
PT genes - for gene therapy with nucleic acid encoding active BLM
PT protein to treat Bloom's syndrome and cancer in general
PT Disclosure; Page -: 51pp; English.
CC This is a mutated BLM gene sequence isolated from an Italian
CC Bloom's Syndrome sufferer designated "113(Dadem)". The substitution of
CC the bp G at the position 3238 of the wild type H1-5' gene to the bp C
CC results in the amino acid Serine at the amino acid position 1055.
CC This is one of the seven unique mutations which were identified in
CC a study of 10 people with Bloom's Syndrome. Based on the various
CC mutations, diagnostic tests for Bloom's Syndrome have been developed
CC which use standard sequence analysis techniques to detect the presence
CC of 2 mutated BLM genes or the absence of a wild-type BLM gene.
CC Note: The present sequence does not appear in the specification; it
CC has been made by modifying the H1-5' wild type BLM sequence which is
CC provided in Figure 2 (T67013).
SQ Sequence 4437 BP; 1465 A; 868 C; 910 G; 1194 T;

Query Match 2.3%; Score 20; DB 1; Length 4437;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 210 tgaagacttataatgatg 229  
|||||  
Db 3076 TGAAGACTTATATGATG 3095

Search completed: May 1, 2000, 18:54:09  
Job time: 18933 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:11:07 ; Search time 181.17 Seconds  
(without alignments)  
563.691 Million cell updates/sec

Title: US-09-215-435-123  
Perfect score: 853  
Sequence: 1 ggaagatggcgagcagctct.....tttgcgaaaaaaaaaaaaa 853

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 214294 seqs, 59861574 residues

Word size : 0

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2.6/prodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2.6/prodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2.6/prodata/1/ina/5C\_COMB.seq.\*  
4: /cgn2.6/prodata/1/ina/5D\_COMB.seq.\*  
5: /cgn2.6/prodata/1/ina/6\_COMB.seq.\*  
6: /cgn2.6/prodata/1/ina/PCU99\_COMB.seq.\*  
7: /cgn2.6/prodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20	2.3	1419	1 US-08-103-739B-1	Sequence 1, Appl1
C 2	20	2.3	1419	2 US-08-474-404-1	Sequence 1, Appl1
C 3	20	2.3	1419	3 US-08-485-845-1	Sequence 1, Appl1
C 4	20	2.3	1419	3 US-08-482-714-1	Sequence 1, Appl1
5	20	2.3	4437	2 US-08-559-303B-72	Sequence 72, Appl1
6	19	2.2	1031	1 US-08-181-271A-106	Sequence 106, App
7	19	2.2	1031	1 US-08-449-315-106	Sequence 106, App
8	19	2.2	1031	1 US-08-444-803-106	Sequence 106, App
9	19	2.2	1031	1 US-08-449-043-106	Sequence 106, App
10	19	2.2	1031	2 US-08-456-265A-106	Sequence 106, App
11	19	2.2	1031	2 US-08-455-416-106	Sequence 106, App
12	19	2.2	1031	2 US-08-455-244-106	Sequence 106, App
13	19	2.2	1031	2 US-08-454-876-106	Sequence 106, App
14	19	2.2	1031	3 US-08-457-364-106	Sequence 106, App
15	19	2.2	1031	3 US-08-456-262-106	Sequence 106, App
16	19	2.2	1031	3 US-08-456-240-106	Sequence 106, App
17	19	2.2	1031	3 US-08-455-736-106	Sequence 106, App
18	19	2.2	1031	4 US-08-971-217-106	Sequence 106, App
19	19	2.2	1369	2 US-08-609-572-3	Sequence 3, Appl1
20	19	2.2	2160	2 US-08-700-749A-8	Sequence 8, Appl1
C 21	19	2.2	4816	2 US-08-592-214A-22	Sequence 22, Appl1
22	18	2.1	1394	4 US-08-068-729-3	Sequence 3, Appl1
23	18	2.1	1617	3 US-08-378-939-9	Sequence 9, Appl1
C 24	17	2.0	20	4 US-08-715-461-5	Sequence 5, Appl1
C 25	17	2.0	32	4 US-08-784-208-3	Sequence 3, Appl1
C 26	17	2.0	165	4 US-08-783-395-3	Sequence 3, Appl1

27	17	2.0	165	4 US-08-924-838-9	Sequence 9, Appl1
C 28	17	2.0	209	2 US-08-510-032A-8	Sequence 8, Appl1
C 29	17	2.0	209	5 US-08-688-514-8	Sequence 8, Appl1
30	17	2.0	319	1 US-08-341-568-5	Sequence 5, Appl1
31	17	2.0	319	3 US-08-911-020-5	Sequence 5, Appl1
32	17	2.0	549	2 US-08-482-142-192	Sequence 192, App
33	17	2.0	549	2 US-08-482-142-194	Sequence 194, App
34	17	2.0	549	2 US-08-482-142-196	Sequence 196, App
35	17	2.0	549	4 US-08-478-572-192	Sequence 192, App
36	17	2.0	549	4 US-08-478-572-194	Sequence 194, App
37	17	2.0	549	4 US-08-478-572-196	Sequence 196, App
38	17	2.0	648	4 US-08-783-395-4	Sequence 4, Appl1
39	17	2.0	834	1 US-07-945-288-1	Sequence 1, Appl1
40	17	2.0	834	1 US-08-462-831-1	Sequence 1, Appl1
41	17	2.0	834	2 US-08-461-441-1	Sequence 1, Appl1
42	17	2.0	834	2 US-08-461-441-1	Sequence 1, Appl1
43	17	2.0	834	2 US-08-482-142-1	Sequence 1, Appl1
44	17	2.0	834	4 US-08-478-572-1	Sequence 1, Appl1
45	17	2.0	834	6 PCT-US93-08518-1	Sequence 1, Appl1

## ALIGNMENTS

RESULT 1  
US-08-103-739B-1/c  
; Sequence 1, Application US/08103739B  
; Patent No. 5478369  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, Marc C.  
; APPLICANT: BEACH, Larry R.  
; APPLICANT: HOWARD, John A.  
; APPLICANT: HUFFMAN, Gary A.  
; TITLE OF INVENTION: DNA Sequences Mediating Male Fertility  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pioneer Hi-Bred International, Inc.  
; STREET: 700 Capital Square, 400 Locust Street  
; CITY: Des Moines  
; STATE: Iowa  
; COUNTRY: U.S.  
; ZIP: 50309  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/103,739B  
; FILING DATE: 02-AUG-1993  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/537,183  
; FILING DATE: 12-JUN-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sweeney, Patricia A.  
; REGISTRATION NUMBER: 32,733  
; REFERENCE/DOCKET NUMBER: 0125R2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (515) 248 4897  
; TELEFAX: (515) 248-4844  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1419 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-103-739B-1

Query Match 2.3%; Score 20; DB 1; Length 1419;  
Best Local Similarity 100.0%; Pred. No. 2.6;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 ctccattggcacagacttct 124  
|||||  
Db 430 CTCATTGGCACAGACTTCT 411

## RESULT 2

US-08-474-404-1/c  
; Sequence 1, Application US/08474404  
; Patent No. 5824524  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, Marc C.  
; APPLICANT: BEACH, Larry R.  
; APPLICANT: HOWARD, John A.  
; APPLICANT: HUFFMAN, Gary A.  
; TITLE OF INVENTION: Nucleotide Sequences Mediating Fertility and Method of Using  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pioneer Hi-Bred International, Inc.  
; STREET: 700 Capital Square, 400 Locust Street  
; CITY: Des Moines  
; STATE: Iowa  
; COUNTRY: U.S.  
; ZIP: 50309  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/474,404  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/103,739  
; FILING DATE: 02-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/537,183  
; FILING DATE: 12-JUN-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sweeney, Patricia A.  
; REGISTRATION NUMBER: 32,733  
; REFERENCE/DOCKET NUMBER: 0125R2R3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (515) 248 4897  
; TELEFAX: (515) 248 4844  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1419 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
US-08-474-404-1

Query Match 2.3%; Score 20; DB 2; Length 1419;

Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 ctccattggcacagacttct 124  
|||||  
Db 430 CTCATTGGCACAGACTTCT 411

## RESULT 3

US-08-485-845-1/c  
; Sequence 1, Application US/08485845  
; Patent No. 5850014  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, Marc C.  
; APPLICANT: BEACH, Larry R.  
; APPLICANT: HOWARD, John A.

; APPLICANT: HUFFMAN, Gary A.  
; TITLE OF INVENTION: Nucleotide Sequences Mediating Fertility and Method of Using  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pioneer Hi-Bred International, Inc.  
; STREET: 700 Capital Square, 400 Locust Street  
; CITY: Des Moines  
; STATE: Iowa  
; COUNTRY: U.S.  
; ZIP: 50309  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,845  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/103,739  
; FILING DATE: 02-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/537,183  
; FILING DATE: 12-JUN-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sweeney, Patricia A.  
; REGISTRATION NUMBER: 32,733  
; REFERENCE/DOCKET NUMBER: 0125R2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (515) 248 4897  
; TELEFAX: (515) 248 4844  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1419 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
US-08-485-845-1

Query Match 2.3%; Score 20; DB 3; Length 1419;

Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 ctccattggcacagacttct 124  
|||||  
Db 430 CTCATTGGCACAGACTTCT 411

## RESULT 4

US-08-482-714-1/c  
; Sequence 1, Application US/08482714  
; Patent No. 5859341  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, Marc C.  
; APPLICANT: BEACH, Larry R.  
; APPLICANT: HOWARD, John A.  
; APPLICANT: HUFFMAN, Gary A.  
; TITLE OF INVENTION: Nucleotide Sequences Mediating Fertility and Method of Using  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pioneer Hi-Bred International, Inc.  
; STREET: 700 Capital Square, 400 Locust Street  
; CITY: Des Moines  
; STATE: Iowa  
; COUNTRY: U.S.  
; ZIP: 50309  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/482.714  
;; FILING DATE: 07-JUN-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/103.739  
;; FILING DATE: 02-AUG-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/537.183  
;; FILING DATE: 12-JUN-1990  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Sweeney, Patricia A.  
;; REGISTRATION NUMBER: 32.733  
;; REFERENCE/DOCKET NUMBER: 0125R2R2  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (515) 248 4897  
;; TELEFAX: (515) 248-4844  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1419 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cdna  
US-08-482-714-1

Query Match 2.3%; Score 20; DB 3; Length 1419;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 ctccattggcacagacttct 124  
|||  
Db 430 CTCATTGGCACAGACTTCT 411

RESULT 5  
US-08-559-303B-72  
; Sequence 72, Application US/08559303B  
; Patent No. 5824501  
; GENERAL INFORMATION:  
; APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA  
; APPLICANT: GRODAN  
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT  
; TITLE OF INVENTION: OF BLOOM'S SYNDROME  
; NUMBER OF SEQUENCES: 78  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN  
; STREET: 90 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10016  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE DISKETTE  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/559.303B  
; FILING DATE: NOVEMBER 15, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ELIZABETH A. BOGOSIAN  
; REGISTRATION NUMBER: 39.911  
; REFERENCE/DOCKET NUMBER: 63475/65  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 697-5995  
; TELEFAX: (212) 286-0854 or 286-0082  
; TELE: TWX 710-581-4766  
; INFORMATION FOR SEQ ID NO: 72:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4437  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: SINGLE

;; TOPOLOGY: LINEAR  
;; MOLECULE TYPE:  
;; DESCRIPTION: OTHER NUCLEIC ACID  
;; HYPOTHETICAL: YES  
;; ANTI-SENSE: NO  
;; FEATURE:  
;; NAME/KEY:  
;; LOCATION:  
;; IDENTIFICATION METHOD:  
;; OTHER INFORMATION:  
US-08-559-303B-72

Query Match 2.3%; Score 20; DB 2; Length 4437;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 tgaagaagcttataatgatg 229  
|||  
Db 3076 TGAAGAAGCTTATAATGATG 3095

RESULT 6  
US-08-181-271A-106  
; Sequence 106, Application US/08181271A  
; Patent No. 5614395  
; GENERAL INFORMATION:  
; APPLICANT: Ryals, John A.  
; APPLICANT: Alexander, Danny C.  
; APPLICANT: Beck, James J.  
; APPLICANT: Duesing, John H.  
; APPLICANT: Friedrich, Leslie B.  
; APPLICANT: Goodman, Robert M.  
; APPLICANT: Harms, Christian  
; APPLICANT: Meins, Jr., Frederick  
; APPLICANT: Montoya, Alice  
; APPLICANT: Moyer, Mary B.  
; APPLICANT: Neuhaus, Jean-Marc  
; APPLICANT: Payne, George B.  
; APPLICANT: Sperison, Christoph  
; APPLICANT: Stinson, Jeffrey R.  
; APPLICANT: Uknes, Scott J.  
; APPLICANT: Ward, Eric R.  
; APPLICANT: Williams, Shericca C.  
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/181.271A  
; FILING DATE: 13-JAN-94  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/093.301  
; FILING DATE: 16-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937.197  
; FILING DATE: 6-NOV-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/678.378  
; FILING DATE: 1-APR-1991  
; PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: US 07/305,566
; FILING DATE: 6-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/165,667
; FILING DATE: 8-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,847
; FILING DATE: 6-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/632,441
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/425,504
; FILING DATE: 20-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,506
; FILING DATE: 6-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/768,122
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/580,431
; FILING DATE: 7-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/368,672
; FILING DATE: 20-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/329,018
; FILING DATE: 24-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/045,957
; FILING DATE: 12-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1031 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-181-271A-106

Query Match 2.2%; Score 19; DB 1; Length 1031;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 835 ttgcgaaaaaataaaaaa 853
Db 1010 TTGCGAAAAAATAAAAAA 1028

RESULT 7
US-08-449-315-106
; Sequence 106: Application US/08449315
; Patent No. 5650505
; GENERAL INFORMATION:
; APPLICANT: Ryals, John A.
; APPLICANT: Alexander, Danny C.
; APPLICANT: Beck, James J.
; APPLICANT: Duesing, John H.
; APPLICANT: Friedrich, Leslie B.
; APPLICANT: Goodman, Robert M.
; APPLICANT: Harms, Christian
; APPLICANT: Meins, Jr., Frederick
; APPLICANT: Montoya, Alice
; APPLICANT: Moyer, Mary B.
; APPLICANT: Neuhaus, Jean-Marc
; APPLICANT: Payne, George B.
; APPLICANT: Sperison, Christoph
; APPLICANT: Stinson, Jeffrey R.
; APPLICANT: Uknes, Scott J.
; APPLICANT: Ward, Eric R.
; APPLICANT: Williams, Shericca C.
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
; DNA SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,315
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/181,271
; FILING DATE: 13-JAN-94
; APPLICATION NUMBER: US 08/093,301
; FILING DATE: 16-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,197
; FILING DATE: 6-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/678,378
; FILING DATE: 1-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/305,566
; FILING DATE: 6-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/165,667
; FILING DATE: 8-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,847
; FILING DATE: 6-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/632,441
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/425,504
; FILING DATE: 20-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,506
; FILING DATE: 6-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/768,122
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/580,431
; FILING DATE: 7-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/368,672
; FILING DATE: 20-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/329,018
; FILING DATE: 24-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/045,957
; FILING DATE: 12-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
```



; REFERENCE/DOCKET NUMBER: S-19825/Pl/CGC 1727  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919)541-8614  
; TELEFAX: (919)541-8689  
; INFORMATION FOR SEQ ID NO: 106:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1031 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-449-315-106

Query Match 2.2%; Score 19; DB 1; Length 1031;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 835 ttgcgaaaaa 853  
|||||  
Db 1010 TTTCGAAAAA 1028

## RESULT 8

US-08-444-803-106  
; Sequence 106, Application US/08444803  
; Patent No. 5654114

## ; GENERAL INFORMATION:

; APPLICANT: Ryals, John A.  
; APPLICANT: Alexander, Danny C.  
; APPLICANT: Beck, James J.  
; APPLICANT: Duesing, John H.  
; APPLICANT: Friedrich, Leslie B.  
; APPLICANT: Goodman, Robert M.  
; APPLICANT: Harms, Christian  
; APPLICANT: Meins, Jr., Frederick  
; APPLICANT: Montoya, Alice  
; APPLICANT: Moyer, Mary B.  
; APPLICANT: Neuhaus, Jean-Marc  
; APPLICANT: Payne, George B.  
; APPLICANT: Sperison, Christoph  
; APPLICANT: Stinson, Jeffrey R.  
; APPLICANT: Uknes, Scott J.  
; APPLICANT: Ward, Eric R.  
; APPLICANT: Williams, Shericca C.  
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
; NUMBER OF SEQUENCES: 106

## ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/444,803  
; FILING DATE: 19-MAY-1995  
; CLASSIFICATION: 536

## ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/181,271  
; FILING DATE: 13-JAN-94  
; APPLICATION NUMBER: US 08/093,301  
; FILING DATE: 16-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,197  
; FILING DATE: 6-NOV-1992  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/678,378  
; FILING DATE: 1-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/305,566  
; FILING DATE: 6-FEB-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/165,667  
; FILING DATE: 8-MAR-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/042,847  
; FILING DATE: 6-APR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/532,441  
; FILING DATE: 21-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/425,504  
; FILING DATE: 20-OCT-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/848,506  
; FILING DATE: 6-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/768,122  
; FILING DATE: 27-SEP-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/580,431  
; FILING DATE: 7-SEP-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/368,672  
; FILING DATE: 20-JUN-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/329,018  
; FILING DATE: 24-MAR-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/045,957  
; FILING DATE: 12-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elmer, James Scott  
; REGISTRATION NUMBER: 36,129  
; REFERENCE/DOCKET NUMBER: S-19825/Pl/CGC 1727  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919)541-8614  
; TELEFAX: (919)541-8689  
; INFORMATION FOR SEQ ID NO: 106:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1031 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-444-803-106

Query Match 2.2%; Score 19; DB 1; Length 1031;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 835 ttgcgaaaaa 853  
|||||

Db 1010 TTTCGAAAAA 1028

## RESULT 9

US-08-449-043-106  
; Sequence 106, Application US/08449043  
; Patent No. 5689044

## ; GENERAL INFORMATION:

; APPLICANT: Ryals, John A.  
; APPLICANT: Alexander, Danny C.  
; APPLICANT: Beck, James J.  
; APPLICANT: Duesing, John H.  
; APPLICANT: Friedrich, Leslie B.  
; APPLICANT: Goodman, Robert M.  
; APPLICANT: Harms, Christian

APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericca C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449,043  
FILING DATE: 24-MAY-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT 1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993

ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1031 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-449-043-106  
Query Match 2.2%; Score 19; DB 1; Length 1031;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 835 ttgcgcaaaaaa 853  
Db 1010 TTGCGAAAAA 1028  
RESULT 10  
US-08-456-265A-106  
Sequence 106, Application US/08456265A  
Patent No. 5767389  
GENERAL INFORMATION:  
APPLICANT: Alexander, Danny C.  
APPLICANT: Ryals, John A.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Stinson, Jeffrey R.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 520 White Plains Road, P.O. Box 2005  
CITY: Tarrytown  
STATE: New York  
COUNTRY: USA  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,265A  
FILING DATE: 31-MAY-95  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/181,271  
FILING DATE: 13-JAN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: S-19825/PI/CGC 1727/DIV10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8587  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1031 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-456-265A-106

Query Match 2.2%; Score 19; DB 2; Length 1031;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 835 ttgcgaataaaataaaaaa 853  
Db 1010 TTGCGAAAAAATAAAAAA 1028

RESULT 11  
US-08-456-416-106  
Sequence 106, Application US/08455416  
Patent No. 5777200  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericca C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,416  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/PI/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:

; LENGTH: 1031 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-455-416-106

Query Match 2.2%; Score 19; DB 2; Length 1031;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 835 ttgcgaaaaaiaaaaaa 853  
|||||iiiiiiiiii  
Db 1010 TTGCGAAAAAIAAAAAA 1028

## RESULT 12

US-08-455-244-106  
; Sequence 106, Application US/08455244  
; Patent No. 5789214

## ; GENERAL INFORMATION:

; APPLICANT: Ryals, John A.  
; APPLICANT: Alexander, Danny C.  
; APPLICANT: Beck, James J.  
; APPLICANT: Duesing, John H.  
; APPLICANT: Friedrich, Leslie B.  
; APPLICANT: Goodman, Robert M.  
; APPLICANT: Harms, Christian  
; APPLICANT: Meins, Jr., Frederick  
; APPLICANT: Montoya, Alice  
; APPLICANT: Moyer, Mary B.  
; APPLICANT: Neuhaus, Jean-Marc  
; APPLICANT: Payne, George B.  
; APPLICANT: Sperison, Christoph  
; APPLICANT: Stinson, Jeffrey R.  
; APPLICANT: Uknes, Scott J.  
; APPLICANT: Ward, Eric R.  
; APPLICANT: Williams, Shericca C.  
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
; NUMBER OF SEQUENCES: 106

## ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/455,244  
; FILING DATE: 31-MAY-1995

## ; CLASSIFICATION:

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/181,271  
; FILING DATE: 13-JAN-94  
; APPLICATION NUMBER: US 08/093,301  
; FILING DATE: 16-JUL-1993

## ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/937,197  
; FILING DATE: 6-NOV-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/678,378  
; FILING DATE: 1-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/305,566  
; FILING DATE: 6-FEB-1989  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/165,667  
; FILING DATE: 8-MAR-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/042,847  
; FILING DATE: 6-APR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/632,441  
; FILING DATE: 21-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/425,504  
; FILING DATE: 20-OCT-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/848,506  
; FILING DATE: 6-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/768,122  
; FILING DATE: 27-SEP-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/580,431  
; FILING DATE: 7-SEP-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/368,672  
; FILING DATE: 20-JUN-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/329,018  
; FILING DATE: 24-MAR-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/045,957  
; FILING DATE: 12-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elmer, James Scott  
; REGISTRATION NUMBER: 36,129  
; REFERENCE/DOCKET NUMBER: S-19825/PI/CGC 1727  
; TELEPHONE: (919)541-8614  
; TELEFAX: (919)541-8689  
; INFORMATION FOR SEQ ID NO: 106:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1031 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-455-244-106

Query Match 2.2%; Score 19; DB 2; Length 1031;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 835 ttgcgaaaaaiaaaaaa 853  
|||||iiiiiiiiii

Db 1010 TTGCGAAAAAIAAAAAA 1028

## RESULT 13

US-08-454-876-106  
; Sequence 106, Application US/08454876  
; Patent No. 5804693

## ; GENERAL INFORMATION:

; APPLICANT: Ryals, John A.  
; APPLICANT: Alexander, Danny C.  
; APPLICANT: Beck, James J.  
; APPLICANT: Duesing, John H.  
; APPLICANT: Friedrich, Leslie B.  
; APPLICANT: Goodman, Robert M.  
; APPLICANT: Harms, Christian  
; APPLICANT: Meins, Jr., Frederick  
; APPLICANT: Montoya, Alice  
; APPLICANT: Moyer, Mary B.  
; APPLICANT: Neuhaus, Jean-Marc  
; APPLICANT: Payne, George B.  
; APPLICANT: Sperison, Christoph

APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericca C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/454,876  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614

TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1031 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-454-876-106  
Query Match 2.28; Score 19; DB 2; Length 1031;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 835 ttgcgaaaaa 853  
|||||  
Db 1010 TTGCGAAAAA 1028  
RESULT 14  
US-08-457-364-106  
Sequence 106, Application US/08457364  
Patent No. 5847258  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericca C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,364  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: US 07/305,566
; FILING DATE: 6-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/165,667
; FILING DATE: 8-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,847
; FILING DATE: 6-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/632,441
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/425,504
; FILING DATE: 20-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,506
; FILING DATE: 6-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/768,122
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/580,431
; FILING DATE: 7-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/368,672
; FILING DATE: 20-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/329,018
; FILING DATE: 24-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/045,957
; FILING DATE: 12-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1031 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-457-364-106

```

```

Query Match 2.28; Score 19; DB 3; Length 1031;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 835 ttgcgaaaaaataaaaaa 853
Db 1010 TTGCGAAAAAATAAAAAA 1028

```

```

RESULT 15
US-08-456-262-106
; Sequence 106, Application US/08456262
; Patent No. 5851766
; GENERAL INFORMATION:
; APPLICANT: Ryals, John A.
; APPLICANT: Alexander, Danny C.
; APPLICANT: Beck, James J.
; APPLICANT: Duesing, John H.
; APPLICANT: Friedrich, Leslie B.
; APPLICANT: Goodman, Robert M.
; APPLICANT: Harms, Christian
; APPLICANT: Meins, Jr., Frederick
; APPLICANT: Montoya, Alice
; APPLICANT: Moyer, Mary B.

```

```

; APPLICANT: Neuhaus, Jean-Marc
; APPLICANT: Payne, George B.
; APPLICANT: Sperison, Christoph
; APPLICANT: Stinson, Jeffrey R.
; APPLICANT: Uknes, Scott J.
; APPLICANT: Ward, Eric R.
; APPLICANT: Williams, Shericca C.
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,262
; FILING DATE: 31-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/181,271
; FILING DATE: 13-JAN-94
; APPLICATION NUMBER: US 08/093,301
; FILING DATE: 16-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,197
; FILING DATE: 6-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/678,378
; FILING DATE: 1-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/305,566
; FILING DATE: 6-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/165,667
; FILING DATE: 8-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,847
; FILING DATE: 6-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/632,441
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/425,504
; FILING DATE: 20-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,506
; FILING DATE: 6-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/768,122
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/580,431
; FILING DATE: 7-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/368,672
; FILING DATE: 20-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/329,018
; FILING DATE: 24-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/045,957
; FILING DATE: 12-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129

```

```

; REFERENCE/DOCKET NUMBER: S-19825/PL/CCG 1727
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1031 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-456-262-106

```

```

Query Match      2.2%; Score 19; DB 3; Length 1031;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 835 ttgcgaaaaaaaaaaaaa 853
   |||||||||||||||
Db 1010 TTGCGAAAAAAAAAAAAA 1028

```

Search completed: May 1, 2000, 16:11:14  
Job time: 18739 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 20:00:14 ; Search time 4088.29 Seconds  
(without alignments)  
-715.488 Million cell updates/sec

Title: US-09-215-435-123  
Perfect score: 853  
Sequence: 1 ggagggatggcgagcagtctt.....tttgcgaaaaaaaaaaaaa 853

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 5265378 seqs, -1714608768 residues

Word size : 0

Total number of hits satisfying chosen parameters: 10530736

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Pending\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/1/pna/PCTUS9\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/pna/US081A\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pna/US081B\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/pna/US081C\_COMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pna/US081D\_COMB.seq.\*
- 8: /cgn2\_6/ptodata/1/pna/US082A\_COMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pna/US082B\_COMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pna/US082C\_COMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pna/US083A\_COMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pna/US083B\_COMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pna/US084A\_COMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pna/US084B\_COMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pna/US084C\_COMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pna/US087A\_COMB.seq.\*
- 19: /cgn2\_6/ptodata/1/pna/US087B\_COMB.seq.\*
- 20: /cgn2\_6/ptodata/1/pna/US087C\_COMB.seq.\*
- 21: /cgn2\_6/ptodata/1/pna/US088A\_COMB.seq.\*
- 22: /cgn2\_6/ptodata/1/pna/US088B\_COMB.seq.\*
- 23: /cgn2\_6/ptodata/1/pna/US088C\_COMB.seq.\*
- 24: /cgn2\_6/ptodata/1/pna/US088D\_COMB.seq.\*
- 25: /cgn2\_6/ptodata/1/pna/US089A\_COMB.seq.\*
- 26: /cgn2\_6/ptodata/1/pna/US089B\_COMB.seq.\*
- 27: /cgn2\_6/ptodata/1/pna/US089C\_COMB.seq.\*
- 28: /cgn2\_6/ptodata/1/pna/US089D\_COMB.seq.\*
- 29: /cgn2\_6/ptodata/1/pna/US089E\_COMB.seq.\*
- 30: /cgn2\_6/ptodata/1/pna/US089F\_COMB.seq.\*
- 31: /cgn2\_6/ptodata/1/pna/US089G\_COMB.seq.\*
- 32: /cgn2\_6/ptodata/1/pna/US089H\_COMB.seq.\*
- 33: /cgn2\_6/ptodata/1/pna/US089I\_COMB.seq.\*
- 34: /cgn2\_6/ptodata/1/pna/US090A\_COMB.seq.\*
- 35: /cgn2\_6/ptodata/1/pna/US090B\_COMB.seq.\*
- 36: /cgn2\_6/ptodata/1/pna/US090C\_COMB.seq.\*
- 37: /cgn2\_6/ptodata/1/pna/US090D\_COMB.seq.\*
- 38: /cgn2\_6/ptodata/1/pna/US090E\_COMB.seq.\*
- 39: /cgn2\_6/ptodata/1/pna/US091A\_COMB.seq.\*
- 40: /cgn2\_6/ptodata/1/pna/US091B\_COMB.seq.\*
- 41: /cgn2\_6/ptodata/1/pna/US091C\_COMB.seq.\*
- 42: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq.\*
- 43: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq.\*

- 44: /cgn2\_6/ptodata/1/pna/US092C\_COMB.seq.\*
- 45: /cgn2\_6/ptodata/1/pna/US092D\_COMB.seq.\*
- 46: /cgn2\_6/ptodata/1/pna/US092E\_COMB.seq.\*
- 47: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq.\*
- 48: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq.\*
- 49: /cgn2\_6/ptodata/1/pna/US093C\_COMB.seq.\*
- 50: /cgn2\_6/ptodata/1/pna/US093D\_COMB.seq.\*
- 51: /cgn2\_6/ptodata/1/pna/US093E\_COMB.seq.\*
- 52: /cgn2\_6/ptodata/1/pna/US093F\_COMB.seq.\*
- 53: /cgn2\_6/ptodata/1/pna/US094A\_COMB.seq.\*
- 54: /cgn2\_6/ptodata/1/pna/US094B\_COMB.seq.\*
- 55: /cgn2\_6/ptodata/1/pna/US094C\_COMB.seq.\*
- 56: /cgn2\_6/ptodata/1/pna/US094D\_COMB.seq.\*
- 57: /cgn2\_6/ptodata/1/pna/US094E\_COMB.seq.\*
- 58: /cgn2\_6/ptodata/1/pna/US094F\_COMB.seq.\*
- 59: /cgn2\_6/ptodata/1/pna/US094G\_COMB.seq.\*
- 60: /cgn2\_6/ptodata/1/pna/US094H\_COMB.seq.\*
- 61: /cgn2\_6/ptodata/1/pna/US094I\_COMB.seq.\*
- 62: /cgn2\_6/ptodata/1/pna/US094J\_COMB.seq.\*
- 63: /cgn2\_6/ptodata/1/pna/US094K\_COMB.seq.\*
- 64: /cgn2\_6/ptodata/1/pna/US094L\_COMB.seq.\*
- 65: /cgn2\_6/ptodata/1/pna/US094M\_COMB.seq.\*
- 66: /cgn2\_6/ptodata/1/pna/US094N\_COMB.seq.\*
- 67: /cgn2\_6/ptodata/1/pna/US094O\_COMB.seq.\*
- 68: /cgn2\_6/ptodata/1/pna/US094P\_COMB.seq.\*
- 69: /cgn2\_6/ptodata/1/pna/US094Q\_COMB.seq.\*
- 70: /cgn2\_6/ptodata/1/pna/US094R\_COMB.seq.\*
- 71: /cgn2\_6/ptodata/1/pna/US094S\_COMB.seq.\*
- 72: /cgn2\_6/ptodata/1/pna/US094T\_COMB.seq.\*
- 73: /cgn2\_6/ptodata/1/pna/US094U\_COMB.seq.\*
- 74: /cgn2\_6/ptodata/1/pna/US094V\_COMB.seq.\*
- 75: /cgn2\_6/ptodata/1/pna/US094W\_COMB.seq.\*
- 76: /cgn2\_6/ptodata/1/pna/US094X\_COMB.seq.\*
- 77: /cgn2\_6/ptodata/1/pna/US094Y\_COMB.seq.\*
- 78: /cgn2\_6/ptodata/1/pna/US094Z\_COMB.seq.\*
- 79: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq.\*
- 80: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq.\*
- 81: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq.\*
- 82: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq.\*
- 83: /cgn2\_6/ptodata/1/pna/US095E\_COMB.seq.\*
- 84: /cgn2\_6/ptodata/1/pna/US095F\_COMB.seq.\*
- 85: /cgn2\_6/ptodata/1/pna/US095G\_COMB.seq.\*
- 86: /cgn2\_6/ptodata/1/pna/US095H\_COMB.seq.\*
- 87: /cgn2\_6/ptodata/1/pna/US095I\_COMB.seq.\*
- 88: /cgn2\_6/ptodata/1/pna/US095J\_COMB.seq.\*
- 89: /cgn2\_6/ptodata/1/pna/US095K\_COMB.seq.\*
- 90: /cgn2\_6/ptodata/1/pna/US095L\_COMB.seq.\*
- 91: /cgn2\_6/ptodata/1/pna/US095M\_COMB.seq.\*
- 92: /cgn2\_6/ptodata/1/pna/US095N\_COMB.seq.\*
- 93: /cgn2\_6/ptodata/1/pna/US095O\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	853	100.0	853	42	US-09-215-435-123
2	469	55.0	2039	54	US-09-471-275-850
3	458	53.7	2090	72	US-60-090-762-79
4	397	46.5	2127	87	US-60-172-360-23786
5	367	43.0	1318	42	US-09-215-435-122
6	349	40.9	482	45	US-09-289-768-28292
7	349	40.9	482	48	US-09-332-782-22673
8	345	40.4	475	43	US-09-234-611-15284
9	345	40.4	475	45	US-09-289-768-19551
10	345	40.4	475	48	US-09-332-782-21546
11	318	37.3	488	43	US-09-234-611-5709
12	310	36.3	393	18	US-08-726-759A-3878
13	310	36.3	393	55	US-60-005-197-3878

- Sequence 123, App  
Sequence 850, App  
Sequence 79, App1  
Sequence 23786, A  
Sequence 122, App  
Sequence 2692, A  
Sequence 22673, A  
Sequence 15284, A  
Sequence 19551, A  
Sequence 21546, A  
Sequence 5709, Ap  
Sequence 3878, Ap  
Sequence 3878, Ap

```
14 301 35.3 426 46 US-09-293-972-4811
15 294 34.5 1661 68 US-60-070-771-1068
16 289 33.9 445 46 US-09-293-972-1724
17 287 33.6 467 50 US-09-354-899-5294
18 287 33.6 467 51 US-09-371-168-2569
19 277 32.5 429 50 US-09-359-067-9846
20 273 32.0 438 92 US-09-489-036-4724
21 273 32.0 1686 42 US-09-215-435-67
22 269 31.5 902 50 US-09-354-899-8261
23 266 31.2 401 48 US-09-371-168-6929
24 266 31.2 401 48 US-09-332-782-18172
25 266 31.2 401 92 US-09-515-694-18172
26 263 30.8 418 53 US-09-431-517-14222
27 250 29.3 261 36 US-09-044-767-2878
28 250 29.3 276 37 US-09-065-511-550
29 250 29.3 276 64 US-60-044-939-550
30 249 29.2 352 92 US-09-489-036-24094
31 248 29.1 359 92 US-09-489-036-22989
32 247 29.0 400 48 US-09-332-782-16139
33 247 29.0 400 92 US-09-515-694-16139
34 246 28.8 377 51 US-09-362-510-56219
35 233 27.3 253 4 US-08-104-507A-6669
36 233 27.3 253 4 US-08-104-507C-6669
37 219 25.7 1662 42 US-09-205-070-1401
38 219 25.7 1662 49 US-09-340-623-1401
39 215 25.2 485 47 US-09-307-649-25
40 213 25.0 282 92 US-09-540-766-15389
41 210 24.6 495 50 US-09-359-067-1290
42 209 24.5 400 92 US-09-489-036-1348
43 207 24.3 244 29 US-08-951-203-3465
44 207 24.3 244 60 US-60-029-803-3465
45 204 23.9 2076 50 US-09-359-922-1595
```

## ALIGNMENTS

## RESULT 1

```
US-09-215-435-123
; Sequence 123, Application US/09215435A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric,
; APPLICANT: Bougueleret Lydie
; TITLE OF INVENTION: Extended cDNAs
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215.435A
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/069,957
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 123
; LENGTH: 853
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..582
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 31..90
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.4
; FEATURE:
; NAME/KEY: polyA_signal
```

```
; LOCATION: 816..821
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 840..853
US-09-215-435-123
```

Query Match 100.0%; Score 853; DB 42; Length 853;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 853; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 ggaggatggcgagcagctggaatgccagaatggataaacgcttttctacagcatttgta 60
Db 1 ggaggatggcgagcagctggaatgccagaatggataaacgcttttctacagcatttgta 60
Qy 61 atgctgtgtgcttagcctcattccaccatctacatgcagcgcctccattggcacagac 120
Db 61 atgctgtgtgcttagcctcattccaccatctacatgcagcgcctccattggcacagac 120
Qy 121 ttctggtatgaatcgaaagtcagttccagttcaagaaaaattccagtgatttgataaaagc 180
Db 121 ttctggtatgaatcgaaagtcagttccagttcaagaaaaattccagtgatttgataaaagc 180
Qy 181 tgggatgaattcattagtaggaagcagatgaaagaaacttataatgatcaccttttcga 240
Db 181 tgggatgaattcattagtaggaagcagatgaaagaaacttataatgatcaccttttcga 240
Qy 241 tacaatggcagctgggattgtggagacggtgtatcacccatacccaaaacatgcattgg 300
Db 241 tacaatggcagctgggattgtggagacggtgtatcacccatacccaaaacatgcattgg 300
Qy 301 tatagccaccagaaaaagacagagtcatttgatgtgtgcacaaaatgtgagtttcaca 360
Db 301 tatagccaccagaaaaagacagagtcatttgatgtgtgcacaaaatgtgagtttcaca 360
Qy 361 ctaactgagcagttcatggagaaaatttgtatcccggaacacacacacacacacacacac 420
Db 361 ctaactgagcagttcatggagaaaatttgtatcccggaacacacacacacacacacacac 420
Qy 421 ctcttaggacctatctttggcgttgcagtccttccattccttttgcagtttgcagtttgc 480
Db 421 ctcttaggacctatctttggcgttgcagtccttccattccttttgcagtttgcagtttgc 480
Qy 481 atgtgctttggggttgcagtccttccattccttgcagtccttccattccttccaccacc 540
Db 481 atgtgctttggggttgcagtccttccattccttgcagtccttccattccttccaccacc 540
Qy 541 attgccacgggcattctccatctccttgcagatcacccatgctgtgaagtcacaggccac 600
Db 541 attgccacgggcattctccatctccttgcagatcacccatgctgtgaagtcacaggccac 600
Qy 601 gaggtgctctgtgtagatgctccagctgaaatcccaagcctaagctcccaactgcagcca 660
Db 601 gaggtgctctgtgtagatgctccagctgaaatcccaagcctaagctcccaactgcagcca 660
Qy 661 acatcattccagccatgtgtggcagccatcctgtgagtcctcagccttaacaaagccttcag 720
Db 661 acatcattccagccatgtgtggcagccatcctgtgagtcctcagccttaacaaagccttcag 720
Qy 721 aggaacttcagccacagctattatcttactacatccttgcagactccttaataaagaacaa 780
Db 721 aggaacttcagccacagctattatcttactacatccttgcagactccttaataaagaacaa 780
Qy 781 ctgctgagcccaatcaacctatggaactgatagataaaataaataatgtgtgttttcgga 840
Db 781 ctgctgagcccaatcaacctatggaactgatagataaaataaataatgtgtgttttcgga 840
Qy 841 aaaaaaaaaa 853
Db 841 aaaaaaaaaa 853
```

RESULT 2

```

301 301  tatatgccaccagaaaaggacagagtgatcatttgatgtgtgcacaaaagtgtgagtttcaca 360
302 302  |||||||
303 303  |||||||
304 304  |||||||
305 305  tatatgccaccagaaaaggacagagtgatcatttgatgtgtgcacaaaagtgtgagtttcaca 414
306 306  |||||||
307 307  |||||||
308 308  |||||||
309 309  |||||||
310 310  |||||||
311 311  |||||||
312 312  |||||||
313 313  |||||||
314 314  |||||||
315 315  |||||||
316 316  |||||||
317 317  |||||||
318 318  |||||||
319 319  |||||||
320 320  |||||||
321 321  |||||||
322 322  |||||||
323 323  |||||||
324 324  |||||||
325 325  |||||||
326 326  |||||||
327 327  |||||||
328 328  |||||||
329 329  |||||||
330 330  |||||||
331 331  |||||||
332 332  |||||||
333 333  |||||||
334 334  |||||||
335 335  |||||||
336 336  |||||||
337 337  |||||||
338 338  |||||||
339 339  |||||||
340 340  |||||||
341 341  |||||||
342 342  |||||||
343 343  |||||||
344 344  |||||||
345 345  |||||||
346 346  |||||||
347 347  |||||||
348 348  |||||||
349 349  |||||||
350 350  |||||||
351 351  |||||||
352 352  |||||||
353 353  |||||||
354 354  |||||||
355 355  |||||||
356 356  |||||||
357 357  |||||||
358 358  |||||||
359 359  |||||||
360 360  |||||||
361 361  ctaactgagcagttcatgagaaaattgtgttgatcccggaacacacaaatagcgggattgat 420
362 362  |||||||
363 363  |||||||
364 364  |||||||
365 365  |||||||
366 366  |||||||
367 367  |||||||
368 368  |||||||
369 369  |||||||
370 370  |||||||
371 371  |||||||
372 372  |||||||
373 373  |||||||
374 374  |||||||
375 375  |||||||
376 376  |||||||
377 377  |||||||
378 378  |||||||
379 379  |||||||
380 380  |||||||
381 381  |||||||
382 382  |||||||
383 383  |||||||
384 384  |||||||
385 385  |||||||
386 386  |||||||
387 387  |||||||
388 388  |||||||
389 389  |||||||
390 390  |||||||
391 391  |||||||
392 392  |||||||
393 393  |||||||
394 394  |||||||
395 395  |||||||
396 396  |||||||
397 397  |||||||
398 398  |||||||
399 399  |||||||
400 400  |||||||
401 401  |||||||
402 402  |||||||
403 403  |||||||
404 404  |||||||
405 405  |||||||
406 406  |||||||
407 407  |||||||
408 408  |||||||
409 409  |||||||
410 410  |||||||
411 411  |||||||
412 412  |||||||
413 413  |||||||
414 414  |||||||
415 415  ctaactgagcagttcatgagaaaattgtgttgatcccggaacacacaaatagcgggattgat 474
416 416  |||||||
417 417  |||||||
418 418  |||||||
419 419  |||||||
420 420  |||||||
421 421  ctccttaggacctatctttggcgttgccagttcctcttaccttttgcagtttaggtttg 480
422 422  |||||||
423 423  |||||||
424 424  |||||||
425 425  |||||||
426 426  |||||||
427 427  |||||||
428 428  |||||||
429 429  |||||||
430 430  |||||||
431 431  |||||||
432 432  |||||||
433 433  |||||||
434 434  |||||||
435 435  |||||||
436 436  |||||||
437 437  |||||||
438 438  |||||||
439 439  |||||||
440 440  |||||||
441 441  |||||||
442 442  |||||||
443 443  |||||||
444 444  |||||||
445 445  |||||||
446 446  |||||||
447 447  |||||||
448 448  |||||||
449 449  |||||||
450 450  |||||||
451 451  |||||||
452 452  |||||||
453 453  |||||||
454 454  |||||||
455 455  |||||||
456 456  |||||||
457 457  |||||||
458 458  |||||||
459 459  |||||||
460 460  |||||||
461 461  |||||||
462 462  |||||||
463 463  |||||||
464 464  |||||||
465 465  |||||||
466 466  |||||||
467 467  |||||||
468 468  |||||||
469 469  |||||||
470 470  |||||||
471 471  |||||||
472 472  |||||||
473 473  |||||||
474 474  |||||||
475 475  ctccttaggacctatctttggcgttgccagttcctcttaccttttgcagtttaggtttg 534
476 476  |||||||
477 477  |||||||
478 478  |||||||
479 479  |||||||
480 480  |||||||
481 481  atgtgctttggggccttgatcggaactttgcttgcaatttgcgcgaagccttatatccacc 540
482 482  |||||||
483 483  |||||||
484 484  |||||||
485 485  |||||||
486 486  |||||||
487 487  |||||||
488 488  |||||||
489 489  |||||||
490 490  |||||||
491 491  |||||||
492 492  |||||||
493 493  |||||||
494 494  |||||||
495 495  |||||||
496 496  |||||||
497 497  |||||||
498 498  |||||||
499 499  |||||||
500 500  |||||||
501 501  |||||||
502 502  |||||||
503 503  |||||||
504 504  |||||||
505 505  |||||||
506 506  |||||||
507 507  |||||||
508 508  |||||||
509 509  |||||||
510 510  |||||||
511 511  |||||||
512 512  |||||||
513 513  |||||||
514 514  |||||||
515 515  |||||||
516 516  |||||||
517 517  |||||||
518 518  |||||||
519 519  |||||||
520 520  |||||||
521 521  |||||||
522 522  |||||||
523 523  |||||||
524 524  |||||||
525 525  |||||||
526 526  |||||||
527 527  |||||||
528 528  |||||||
529 529  |||||||
530 530  |||||||
531 531  |||||||
532 532  |||||||
533 533  |||||||
534 534  |||||||
535 535  atgtgctttggggccttgatcggaactttgcttgcaatttgcgcgaagccttatatccacc 594
536 536  |||||||
537 537  |||||||
538 538  |||||||
539 539  |||||||
540 540  |||||||
541 541  attgccacgggcattctccatctcttgcag 571
542 542  |||||||
543 543  |||||||
544 544  |||||||
545 545  |||||||
546 546  |||||||
547 547  |||||||
548 548  |||||||
549 549  |||||||
550 550  |||||||
551 551  |||||||
552 552  |||||||
553 553  |||||||
554 554  |||||||
555 555  |||||||
556 556  |||||||
557 557  |||||||
558 558  |||||||
559 559  |||||||
560 560  |||||||
561 561  |||||||
562 562  |||||||
563 563  |||||||
564 564  |||||||
565 565  |||||||
566 566  |||||||
567 567  |||||||
568 568  |||||||
569 569  |||||||
570 570  |||||||
571 571  |||||||
572 572  |||||||
573 573  |||||||
574 574  |||||||
575 575  |||||||
576 576  |||||||
577 577  |||||||
578 578  |||||||
579 579  |||||||
580 580  |||||||
581 581  |||||||
582 582  |||||||
583 583  |||||||
584 584  |||||||
585 585  |||||||
586 586  |||||||
587 587  |||||||
588 588  |||||||
589 589  |||||||
590 590  |||||||
591 591  |||||||
592 592  |||||||
593 593  |||||||
594 594  |||||||
595 595  attgccacgggcattctccatctcttgcag 625
596 596  |||||||
597 597  |||||||
598 598  |||||||
599 599  |||||||
600 600  |||||||
601 601  |||||||
602 602  |||||||
603 603  |||||||
604 604  |||||||
605 605  |||||||
606 606  |||||||
607 607  |||||||
608 608  |||||||
609 609  |||||||
610 610  |||||||
611 611  |||||||
612 612  |||||||
613 613  |||||||
614 614  |||||||
615 615  |||||||
616 616  |||||||
617 617  |||||||
618 618  |||||||
```

```

Query Match      53.74; Score 458; DB 72; Length 2090;
Best Local Similarity 99.68; Pred. No. 6.9e-197;
Matches 558; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 12 gacgagctctgaatgcgaagaatggataaaccttttctctacagcaattgttaattgcttctgt 71

```

```

|||||
Db 121 GAGCAGCTGAAATGCGAATGATGAACCGTTTGTGTACAGCATTTGTAATGCTGTGT 180
|||||
Qy 72 gtttagcctcattccaccatctacatggcagcctccattggcacagactttggtatga 131
|||||
Db 181 GCTTAGCCTCAATTCACCATCTACATGCGAGCCTCCATTGGCACAGACTTCTGTTATGA 240
|||||
Qy 132 atatcgaagtcagttccagtaaaatccagtgatttggaataaaagcatctggatgaatt 191
|||||
Db 241 ATATCGAAGTCCAGTTCAGAAATTCAGTGAATTTGAATAAAGCATCTGGATGAAT 300
|||||
Qy 192 cattagtgaagcagatgaagacttatatgatgcaccttttccgatacaaatggcac 251
|||||
Db 301 CATTAGTGATGAGCAGATGAAAGACTTATATGATGACATTTTCGATACAAATGGCAC 360
|||||
Qy 252 agtgggatttggagacggtgtatcccatatcccaatacccaaaacatgcattgtagcccaac 311
|||||
Db 361 AGTGGGATTGTGGAGACGGTGTATCACCATACCATACCAAAACATGTTGTATAGCCACC 420
|||||
Qy 312 agaaagacagagatcttggatggtgcacaaatgtgagtttcacactaaactagca 371
|||||
Db 421 AGAAGGACAGAGTCATTGATGTGTGCACAAATGTGTGAGTTTCACACTAACATGAGCA 480
|||||
Qy 372 gttcattggagaaattgttgatccccggaaccacacaaatagcgggattgatctccttaggac 431
|||||
Db 481 GTTCATGAGAAATTTGTATCCCGGAAACACAAATAGCGGATTGATCTCCTTAGGAC 540
|||||
Qy 432 ctatcttggcgttgcagttccttttaccttttggattgagttttaggtttagttgctttgg 491
|||||
Db 541 CTATCTTTGGCGTGGCAGTTCCTTTTACCTTTTGTGAGTTTAGTGTGCTTTGG 600
|||||
Qy 492 ggtttgacgagcttctgtctgtcatttgcgaagcttatatccacacattgccacggg 551
|||||
Db 601 GGCTTTGATGGAGATTGTGCTTGCATTTGCCGAGCTTATATCCACCATTCGCCAGGG 660
|||||
Qy 552 cattctccatctccttgag 571
|||||
Db 661 CATTCTCCATCTCCTTGAC 680
|||||

RESULT 4
US-60-172-360-23786
; Sequence 23786, Application US/60172360
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Polynucleotide Sequence Polymorphisms Using
; FILE REFERENCE: GX-0007 P
; CURRENT APPLICATION NUMBER: US/60/172,360
; CURRENT FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 29838
; SOFTWARE: PERL Program
; SEQ ID NO 23786
; LENGTH: 2127
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 403627.7
US-60-172-360-23786

Query Match 46.5%; Score 397; DB 87; Length 2127;
Best Local Similarity 99.6%; Pred. No. 2.7e-169;
Matches 497; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 73 cttagcctcattccaccatctacatggcagcctccattggcacagacttctggtatga 132
Db 228 cttagcctcattccaccatctacatggcagcctccattggcacagacttctggtatga 287
Qy 133 tatcgaagtcagttccaagaaattccagtgatttgaataaaagcatctggatgaattc 192

```

```

|||||
Db 288 tatcgaagtcagttccaagaaattccagtgatttgaataaaagcatctgggatgaattc 347
|||||
Qy 193 attagtgaagcagatgaagaaacttataatgatgcaccttttcgatacaaatggcac 252
|||||
Db 348 attagtgaagcagatgaagaaacttataatgatgcaccttttcgatacaaatggcac 407
|||||
Qy 253 gtgggatttggagacggtgtatcccatatcccaaaacatgcatttggtagccccacca 312
|||||
Db 408 gtgggatttggagacggtgtatcccatatcccaaaacatgcatttggtagccccacca 467
|||||
Qy 313 gaaagacagagtcatttgcatttgcacaaatgtgagtttcacactaaactagcag 372
|||||
Db 468 gaaagacagagtcatttgcatttgcacaaatgtgagtttcacactaaactagcag 527
|||||
Qy 373 ttcattggagaaatttgcatttgcacaaatgtgagtttcacactaaactagcag 432
|||||
Db 528 ttcattggagaaatttgcatttgcacaaatgtgagtttcacactaaactagcag 587
|||||
Qy 433 tatcttggcgttgcagttccttttaccttttgcagtttgcagtttgcagtttgcagtttgcag 492
|||||
Db 588 tatcttggcgttgcagttccttttaccttttgcagtttgcagtttgcagtttgcagtttgcag 647
|||||
Qy 493 gctttgacgagcttgcagtttgcagtttgcagtttgcagtttgcagtttgcagtttgcagtttgcag 552
|||||
Db 648 gctttgacgagcttgcagtttgcagtttgcagtttgcagtttgcagtttgcagtttgcagtttgcag 707
|||||
Qy 553 attctccatctccttgag 571
|||||
Db 708 attctccatctccttgag 726
|||||

```

```

RESULT 5
US-09-215-435-122
; Sequence 122, Application US/09215435A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric,
; APPLICANT: Bouqueleret Lydie
; TITLE OF INVENTION: Extended cDNAs
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215,435A
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/069,957
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 122
; LENGTH: 1318
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..560
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 31..90
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.4
; OTHER INFORMATION: seq AFVIAVLISLIST/IY
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1288..1293
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 1307..1318
; US-09-215-435-122

```

```
Query Match          43.0%; Score 367; DB 42; Length 1318;
Best Local Similarity 99.3%; Pred. No. 1e-155;
Matches 567; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggagatggcgagcagctgtaatgccagaatggataacgcttttgcacagcatttga 60
D 1 ggagatggcgagcagctgtaatgccagaatggataacgcttttgcacagcatttga 60
QY 61 attgcttgcttagctcattccaccatctacatggcagcctccattggcacaac 120
D 61 attgcttgcttagctcattccaccatctacatggcagcctccattggcacaac 120
QY 121 ttctggtatgaatcgaagtcagcttcaagaaaaatccagtgatttgaataaaagcatc 180
D 121 ttctggtatgaatcgaagtcagcttcaagaaaaatccagtgatttgaataaaagcatc 180
QY 181 tgggatgaattcattagtgatgaagcagatgaagagacttataatgatgcaccttttga 240
D 181 tgggatgaattcattagtgatgaagcagatgaagagacttataatgatgcaccttttga 240
QY 241 tacaatggcacagtggtggtgagacggtgtatcacaccatcccaaaaacatgcattgg 300
D 241 tacaatggcacagtggtggtgagacggtgtatcacaccatcccaaaaacatgcattgg 300
QY 301 tatagccaccagaaaggacagagcatttgatggtgcacaaaatgtgtgagtttcaca 360
D 301 tatagccaccagaaaggacagagcatttgatggtgcacaaaatgtgtgagtttcaca 360
QY 361 ctcaactgagcagttcattgagaaaattgtgtatcccggaacacaaatagcgggattgat 420
D 361 ctcaactgagcagttcattgagaaaattgtgtatcccggaacacaaatagcgggattgat 420
QY 421 ctctctaggacattctttggcggtgcagttcctttacccttttgcagtttaggttg 480
D 421 ctctctaggacattctttggcggtgcagttcctttacccttttgcagtttaggttg 480
QY 481 attgctttggggcctttgacggaactttgtcttgcaatttgcgcgaagcttatccccacc 540
D 481 attgctttggggcctttgacggaactttgtcttgcaatttgcgcgaagcttatccccacc 540
QY 541 attgcaacgggcattctccatctccttgcaag 571
D 541 attgcaacgggcattctccatctccttgcaag 571

RESULT 6
US-09-289-768-26292
; Sequence 26292, Application US/09289768
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-765
; CURRENT APPLICATION NUMBER: US/09/289,768
; CURRENT FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 39996
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 26292
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(482)
; OTHER INFORMATION: n = A,T,C or G
US-09-289-768-26292

Query Match          40.9%; Score 349; DB 45; Length 482;
Best Local Similarity 99.6%; Pred. No. 1.4e-147;
Matches 449; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 12 gagcagctcgaatgccagaatggataacgcttttgcacagcatttgaattgcttgt 71
D 12 gagcagctcgaatgccagaatggataacgcttttgcacagcatttgaattgcttgt 71
QY 72 gcttagcctcatttccaccatctacatggcagcctccattggcagacacttctggtatga 131
D 72 gcttagcctcatttccaccatctacatggcagcctccattggcagacacttctggtatga 131
QY 132 atatcgaaagtcagttcaagaaaaatccagtgatttgaataaaagcatctctggatgaatt 191
D 132 atatcgaaagtcagttcaagaaaaatccagtgatttgaataaaagcatctctggatgaatt 191
QY 152 atatcgaaagtcagttcaagaaaaatccagtgatttgaataaaagcatctctggatgaatt 211
D 152 atatcgaaagtcagttcaagaaaaatccagtgatttgaataaaagcatctctggatgaatt 211
QY 192 cattagtatgaagcagatgaagaaacttataatgatgcaccttttcgatacaatggcacc 251
D 192 cattagtatgaagcagatgaagaaacttataatgatgcaccttttcgatacaatggcacc 251
QY 212 cattagtatgagcgagatgaagaaacttataatgatgcaccttttcgatacaatggcacc 271
D 212 cattagtatgagcgagatgaagaaacttataatgatgcaccttttcgatacaatggcacc 271
QY 252 agtgggatttggagacggtgtatcacaccatcccaaaaacatgcattggtatagccacc 311
D 252 agtgggatttggagacggtgtatcacaccatcccaaaaacatgcattggtatagccacc 311
QY 272 agtgggatttggagacggtgtatcacaccatcccaaaaacatgcattggtatagccacc 331
D 272 agtgggatttggagacggtgtatcacaccatcccaaaaacatgcattggtatagccacc 331
QY 312 agaaaggacagagtcatttgcgtggtgcacaaaatgtgtgagtttcacactaaactgagca 371
D 312 agaaaggacagagtcatttgcgtggtgcacaaaatgtgtgagtttcacactaaactgagca 371
QY 332 agaaaggacagagtcatttgcgtggtgcacaaaatgtgtgagtttcacactaaactgagca 391
D 332 agaaaggacagagtcatttgcgtggtgcacaaaatgtgtgagtttcacactaaactgagca 391
QY 372 gtctcatggagaaatttgcgtggtgcacaaaatgtgtgagtttcacactaaactgagca 431
D 372 gtctcatggagaaatttgcgtggtgcacaaaatgtgtgagtttcacactaaactgagca 431
QY 392 gtctcatggagaaatttgcgtggtgcacaaaatgtgtgagtttcacactaaactgagca 451
D 392 gtctcatggagaaatttgcgtggtgcacaaaatgtgtgagtttcacactaaactgagca 451
QY 432 ctatctttggcggtgcagttcctttttacct 462
D 432 ctatctttggcggtgcagttcctttttacct 462
QY 452 ctatctttggcggtgcagttcctttttacct 482
D 452 ctatctttggcggtgcagttcctttttacct 482

RESULT 7
US-09-332-782-22673
; Sequence 22673, Application US/09332782A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/332,782A
; CURRENT FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22673
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(482)
; OTHER INFORMATION: n = A,T,C or G
US-09-332-782-22673

Query Match          40.9%; Score 349; DB 48; Length 482;
Best Local Similarity 99.6%; Pred. No. 1.4e-147;
Matches 449; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 12 gagcagctcgaatgccagaatggataacgcttttgcacagcatttgaattgcttgt 71
D 12 gagcagctcgaatgccagaatggataacgcttttgcacagcatttgaattgcttgt 71
QY 72 gcttagcctcatttccaccatctacatggcagcctccattggcagacacttctggtatga 131
D 72 gcttagcctcatttccaccatctacatggcagcctccattggcagacacttctggtatga 131
QY 132 atatcgaaagtcagttcaagaaaaatccagtgatttgaataaaagcatctctggatgaatt 191
D 132 atatcgaaagtcagttcaagaaaaatccagtgatttgaataaaagcatctctggatgaatt 191
QY 152 atatcgaaagtcagttcaagaaaaatccagtgatttgaataaaagcatctctggatgaatt 211
D 152 atatcgaaagtcagttcaagaaaaatccagtgatttgaataaaagcatctctggatgaatt 211
```

Qy	192	cattagtgtgaagcagatgaagaagacttatataatgtatgcacotcttttcgtatacaatggcac	251
Db	212	cattagtgtgagcgagatgaagaacttataatgtatgcacotcttttcgtatacaatggcac	271
Qy	252	agtgggattgtgagacggtgtatcaccatacccaaaaacatgcattggtatagccacc	311
Db	272	agtgggattgtgagacggtgtatcaccatacccaaaaacatgcattggtatagccacc	331
Qy	312	agaaagcacagatgcatttgcgtgcacaaaatgtgtagtttcacactaactgagca	371
Db	332	agaaagcacagatgcatttgcgtgcacaaaatgtgtagtttcacactaactgagca	391
Qy	372	gttcattgagaaaattgttgcacccggaaaaccacatagcgggattgattctccttaggac	431
Db	392	gttcattgagaaaattgttgcacccggaaaaccacatagcgggattgattctccttaggac	451
Qy	432	ctatctttggcgttgcagttcctctttacct	462
Db	452	ctatctttggcgttgcagttcctctttacct	482
RESULT 8			
US-09-234-611-15284			
; Sequence 15284, Application US/09234611			
; GENERAL INFORMATION:			
; APPLICANT: Hyseq, Inc.			
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED			
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES			
; FILE REFERENCE: 20411-757			
; CURRENT APPLICATION NUMBER: US/09/234,611			
; CURRENT FILING DATE: 1999-01-21			
; NUMBER OF SEQ ID NOS: 21025			
; SOFTWARE: FastSEQ for Windows Version 3.0			
; SEQ ID NO 15284			
; LENGTH: 475			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: misc_feature			
; LOCATION: (1)..(475)			
; OTHER INFORMATION: n = A,T,C or G			
US-09-234-611-15284			

Query Match 40.4%; Score 345; DB 48; Length 475;  
Best Local Similarity 100.0%; Pred. No. 9e-146;  
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 ttctcgatacaatgacacagtgagtgattgtggagacggtgtatcacatcccccaaaacat 293  
DB 131 ttctcgatacaatgacacagtgagtgattgtggagacggtgtatcacatcccccaaaacat 190  
QY 294 gcattggtatagccaccagaagacagagtcattgtggtgcacaaaatgtgtgag 353  
DB 191 gcattggtatagccaccagaagacagagtcattgtggtgcacaaaatgtgtgag 250  
QY 354 ttccacactaatgacagtcattgcattgagaaatttgttgcacccggaacacacataagcgg 413  
DB 251 ttccacactaatgacagtcattgcattgagaaatttgttgcacccggaacacacataagcgg 310  
QY 414 gattgatcccttagacacatcttttggcgttgccagttccctttaccttttgcagttt 473  
DB 311 gattgatcccttagacacatcttttggcgttgccagttccctttaccttttgcagttt 370  
QY 474 aggtttgatgtcttggcgttgcagtcattgtggttcattgcttcattgcagagcttata 533  
DB 371 aggtttgatgtcttggcgttgcagtcattgtggttcattgcttcattgcagagcttata 430  
QY 534 tccaccattgcacgggcatctccatctccttcagataccat 578  
DB 431 tccaccattgcacgggcatctccatctccttcagataccat 475

## RESULT 11

US-09-234-611-5709  
; Sequence 5709, Application US/09234611  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-757  
; CURRENT APPLICATION NUMBER: US/09/234, 611  
; CURRENT FILING DATE: 1999-01-21  
; NUMBER OF SEQ ID NOS: 21025  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5709  
; LENGTH: 488  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc\_feature  
; FEATURE:  
; LOCATION: (1)...(488)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-234-611-5709

Query Match 37.3%; Score 318; DB 43; Length 488;  
Best Local Similarity 100.0%; Pred. No. 1.5e-133;  
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 ttctcgatacaatgacacagtgagtgattgtggagacggtgtatcacatcccccaaaacat 293  
DB 170 ttctcgatacaatgacacagtgagtgattgtggagacggtgtatcacatcccccaaaacat 229  
QY 294 gcattggtatagccaccagaagacagagtcattgtggttcacaaaatgtgtgag 353  
DB 230 gcattggtatagccaccagaagacagagtcattgtggttcacaaaatgtgtgag 289  
QY 354 ttccacactaatgacagtcattgcattgagaaatttgttgcacccggaacacacataagcgg 413  
DB 290 ttccacactaatgacagtcattgcattgagaaatttgttgcacccggaacacacataagcgg 349  
QY 414 gattgatcccttagacacatcttttggcgttgccagttccctttaccttttgcagttt 473  
DB 350 gattgatcccttagacacatcttttggcgttgccagttccctttaccttttgcagttt 409

QY 474 aggtttgatgtcttggcgttgcagtcattgtggttcattgcttcattgcagagcttata 533  
DB 410 aggtttgatgtcttggcgttgcagtcattgtggttcattgcttcattgcagagcttata 469  
QY 534 tccaccattgcacggg 551  
DB 470 tccaccattgcacggg 487

## RESULT 12

US-08-726-759A-3878  
; Sequence 3878, Application US/08726759A  
; GENERAL INFORMATION:  
; APPLICANT: Stuart, Susan G.  
; APPLICANT: Gooding, Douglas H.  
; APPLICANT: Lane, John C.  
; APPLICANT: Deleane, Angelo M.  
; APPLICANT: Snable, James I.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES  
; TITLE OF INVENTION: DERIVED FROM HUMAN HEART  
; NUMBER OF SEQUENCES: 5395  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/726,759A  
; FILING DATE: Filed Herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/005,197  
; FILING DATE: October 10, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CERRONE, MICHAEL C.  
; REGISTRATION NUMBER: 39132  
; REFERENCE/DOCKET NUMBER: PD-0053 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 855-0555  
; TELEFAX: (415) 845-4166  
; INFORMATION FOR SEQ ID NO: 3878:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 393 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; IMMEDIATE SOURCE:  
; CLONE: 349967  
US-08-726-759A-3878

Query Match 36.3%; Score 310; DB 18; Length 393;  
Best Local Similarity 100.0%; Pred. No. 6.1e-130;  
Matches 310; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 tggagacggtgtatcacatcccccaaaacatgcttggatagccaccagaagagaca 321  
DB 83 TGGAGACGGTGTATCACCATACCCAAAACATGCTTGGTATAGCCACCAAGGACA 142  
QY 322 gagtcattgtatggttcacaaaatgtgtgagtttcacactaactgagcagttcatgag 381  
DB 143 GAGTCATTGTATGCTCACAATAATGTGTAGTTTCACACTAAGTGCAGTTCATGGAG 202  
QY 382 aaattgttgatcccggaacacacataagcgggattgcttctcttagacctatttgg 441





Job time: 22005 sec

```
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Au-Young, Janice
; APPLICANT: Guegler, Karl J.
; APPLICANT: Raughn, Mariah R.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: MICROARRAY FOR THE DETECTION OF
; TITLE OF INVENTION: EXPRESSION OF GENES CODING FOR SIGNAL SEQUENCE CONTAINING POLY
; NUMBER OF SEQUENCES: 1157
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/070,771
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0004 P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 1068:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1661 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TLYMN0T04
; CLONE: 2930673
; US-60-070-771-1068
```

```
Query Match      34.5%; Score 294; DB 68; Length 1661;
Best Local Similarity 100.0%; Pred. No. 1.le-122;
Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 ccatacccaaaacatgcatgttgtagccaccagaaaggacagagtcatttgatgtgg 337
Db 391 CCATACCCCAAAACATGCTATGCTATAGCCACCACAGAAAGGACAGAGTCATTGTGATGG 450
QY 338 tcacaaaatgtgtgagtttcacactaacctgagcagttcatgagaaaatttgtgatcccg 397
Db 451 TCACAAAATGTGTGAGTTTCACACTAACTGACAGCTTCATGGAGAAATTTGTGATCCCG 510
QY 398 gaacacacaatagcgggattgctcctcttaggacatctcttggcggttgcagttcccttt 457
Db 511 GAACACACAAATAGCGGGATTGATCTCCTTAGGACCTATCTTTGGCCTTGCCAGTTCCCTTT 570
QY 458 taccttttgtgagtttagttgtgctttggggtttgatcggaactttgtgcttga 517
Db 571 TACCTTTTGTGAGTTTAGTTGTGCTTTGGGCTTTGATCGGACTTTTGTGCTTGA 630
QY 518 ttgccaagagcttatatccacacattgccacgggaattctccatctccttgcag 571
Db 631 TTTGCCGAAGCTTATATCCACATTGCCACGGGCAATCTCCATCTCCTTGCAG 684
```

Search completed: May 1, 2000, 20:00:21

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 14:44:51 ; Search time 3022.95 Seconds  
(without alignments)  
1065.397 Million cell updates/sec

Title: US-09-215-435-123  
Perfect score: 853  
Sequence: 1 ggaggatggcgagcagtct.....tttgcgaaaaaaaaaaaaa 853

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4538634 seqs, 1887831982 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :		EST:*
1:	em_est1:	*
2:	em_est2:	*
3:	em_est3:	*
4:	em_est4:	*
5:	em_est5:	*
6:	em_est6:	*
7:	em_est7:	*
8:	em_est8:	*
9:	em_est9:	*
10:	em_est10:	*
11:	em_est11:	*
12:	em_est12:	*
13:	em_est13:	*
14:	em_est14:	*
15:	em_est15:	*
16:	em_est16:	*
17:	em_est17:	*
18:	em_est18:	*
19:	em_est19:	*
20:	gb_est1:	*
21:	gb_est2:	*
22:	gb_est3:	*
23:	gb_est4:	*
24:	gb_est5:	*
25:	gb_est6:	*
26:	gb_est7:	*
27:	gb_est8:	*
28:	gb_est9:	*
29:	gb_est10:	*
30:	gb_est11:	*
31:	gb_est12:	*
32:	gb_est13:	*
33:	gb_est14:	*
34:	gb_est15:	*
35:	gb_est16:	*
36:	gb_est17:	*
37:	gb_est18:	*
38:	gb_est19:	*
39:	gb_est20:	*
40:	gb_est21:	*
41:	gb_est22:	*
42:	gb_est23:	*
43:	gb_est24:	*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result	Query		
No.	Score	Match	Length DB ID
-----			
c 1	378	44.3	388 43 A1183646
-----			
			Description
			A1183646 qf30f09.x

DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html		Insert Length: 531    Std Error: 0.00 Seq primer: -400P from Gibco High quality sequence stop: 376. Location/Qualifiers 1..388 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1751561" /clone_lib="Soares_testis_NHT" /sex="male" /lab_host="DH10B" /note="vector: pT7T3-Pac (Pharmacia) with a modified polylinker; site_1: Not I; site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTATCCCAATCTGAAGTGGGAGCGGCCCAATTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."	
BASE COUNT    110 a    82 c    98 g    98 t		ORIGIN	
Query Match    44.3%;    Score 378;    DB 43;    Length 388; Best Local Similarity    100.0%;    Pred. No. 7.3e-165; Matches 378;    Conservative    0;    Mismatches    0;    Indels    0;    Gaps    0;			
QY	460	cttttggagtttaggttgatgtgcttgggcttggatcggaacttggcttgcatt	519
Db	388	CCTTTTGTGAGTTTAGGTTGATGCTTTGGGGCTTTGTATCGGACTTTGCTTGCATT	329
QY	520	tgcgaagcttatatcccacattgccacggcattctccatctccttcgagataccatg	579
Db	328	TGCCGAAGCTTATATCCACCATTGCCACGGGCAATCTCCATCTCCTTCAGATACCATG	269
QY	580	ctgtgaagtcaggccacatggaggtgctctgtgtagatgctccagctgaatcccaacg	639
Db	268	CTGTGAAGTCCAGGCCACATGGAGGTGCTCTGTGTAGATGCTCCAGCTGAATCCCAAGC	209
QY	640	taagctcccaactgacagcacaacatcatttccagccatgtgtggagccatccttgatgt	699
Db	208	TAAGCTCCCAACTGACAGCCACATCATTTCCAGCCATGTGTGGAGGCCATCCTGGATGT	149
QY	700	ccagccttaaacagccttcagaggacttcagccacagctattctattctatacatccttgt	759
Db	148	CCAGCCTTTAAACAGCCTTCAGAGGACTTCAGCCACAGCTATTATCTTACTACATCCTTGT	89
QY	760	gagactcttaataaagaaccaactagctgagcccaatcaacctatgaactgatagaata	819
Db	88	GAGACTCTTAATAAAGAACCAACTAGCTGAGCCCAATCAACCTATGAAGACTGATAGAATA	29
QY	820	aaatgaattgtgtttg 837	
Db	28	AAATGAATTGTTGTTTG 11	
RESULT	2		
LOCUS	AA769827/c		
DEFINITION	ah81c02.s1 Soares_testis_NHT Homo sapiens cDNA clone 1322018 3',		
ACCESSION	AA769827		
VERSION	AA769827.1	GI:2821065	
KEYWORDS	EST.		
SOURCE	human.		

QY	829	gtgtgttttg 837 
Db	44	gttgttttg 36 
RESULT 3		
AA040168/c		
LOCUS		
DEFINITION		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
COMMENT		
SOURCE		
FEATURES		
source		
BASE COUNT		
ORIGIN		
Query Match		
Best Local Similarity		
Matches		
QY		
Db		
QY		
Db		
QY		
Db		
QY		



Site\_2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo(dT) primer [5',  
AAGTGAAGATTCCGCGCCGCTTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by M. Fatima Bonaldo."

BASE COUNT 86 a 62 c 90 g 89 t 4 others

Query Match 31.8%; Score 271; DB 29; Length 331;  
Best Local Similarity 100.0%; Pred. No. 2.7e-115;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 545 ccacgggcatctcattctcagataccatgctgtaagtcagccagccacatggag 604  
|||||  
Db 300 CCACGGGCAATCTCCATCTCTTGAGATACCATGCTGTGAAGTCCAGCCACATGGAGG 241  
|||||

Qy 605 tgcctgtgtagatgctcagctgaatcccaagctaaagctcccaactgacagccaacat 664  
|||||  
Db 240 TGTCTGTGTAGATGCTCAGCTGAATCCCAAGCTAAAGCTCCCAACTGACAGCCAAACAT 181  
|||||

Qy 665 catttccagccatgctgggagccatcctgtagtgcagcccttaacaagccttcagagga 724  
|||||  
Db 180 CATTTCCAGCCATGTGTGGAGCCATCTGGATGTCCAGCCCTTAACAAGCCTTCAGAGGA 121  
|||||

Qy 725 ctccagccacagctattcttactacatcctgtgagactctaataaagaaccaactag 784  
|||||  
Db 120 CTTGACCCACAGCTATTATCTTACTACATCTTGTGAGACTCTAATAAGAACCAACTAG 61  
|||||

Qy 785 ctgagcccaatcaacctatggaactgata 815  
|||||  
Db 60 CTGAGCCCAATCAACTATGGAACGTAGATA 30  
|||||

RESULT 6  
LOCUS A1240445 402 bp mRNA EST 04-NOV-1998  
DEFINITION qu41c10.x1 NCI\_CGAP\_Lym5 Homo sapiens cDNA clone IMAGE:1967346 3',  
mRNA sequence.

ACCESSION A1240445  
VERSION A1240445.1 GI:3835842  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 402)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Unknown library type  
Seq primer: -40UP from Gibco  
High quality sequence stop: 381.  
Location/Qualifiers  
1..402  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1967346"  
/clone.lib="NCI\_CGAP\_Lym5"  
/tissue\_type="follicular lymphoma"  
/lab\_host="SOUR (Stratagene, kanamycin resistant)"  
/note="Organ: lymph node; Vector: pBluescript SK-; Site:1:  
EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Average insert size 1.2 kb. Non-amplified  
library. -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'  
adaptor sequence: 5' CTCAGTTTATTTTTTTTTTTT 3"

BASE COUNT 104 a 79 c 104 g 115 t

ORIGIN

Query Match 30.6%; Score 261; DB 43; Length 402;  
Best Local Similarity 100.0%; Pred. No. 1.1e-110;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 569 cagataccatgctgtaagtcagccacatgaggggtcctgtgtagatgctccagctg 628  
|||||  
Db 261 CAGATACCATGCTGTGAAGTCCAGCCACATGGAGGTCTCTGTGTAGATGCTCCAGCTG 202  
|||||

Qy 629 aaatcccaagctaaagtcctcccaactgaagccacatcatttccagccatgctgggagcc 688  
|||||  
Db 201 AAATCCCAAGCTAAGCTCCCAACTGACAGCCACATCATTTCCAGCCATGTGTGGAGCC 142  
|||||

Qy 689 atcctgtagtccagccttaacaagccttcagagcattcagccacagctattatcttac 748  
|||||  
Db 141 ATCTGTGATGCCAGCCTTAACAAGCCTTCAGAGGACTTCAGCCACAGCTATTATCTTAC 82  
|||||

Qy 749 tacatctgtgagactctaataaagaaccaactagctgagcccaatcaacctatggaac 808  
|||||  
Db 81 TACATCTGTGAGACTCTAATAAAGAACCAACTAGCTGAGCCCAATCAACCTATGGAAC 22  
|||||

Qy 809 tgatagaaaataaatgaattg 829  
|||||  
Db 21 TGATAGAAAATAAATGAATTG 1  
|||||

RESULT 7  
LOCUS T35997 253 bp mRNA EST 06-SEP-1995  
DEFINITION EST95501 Human Testis Homo sapiens cDNA 5' end similar to None,  
mRNA sequence.

ACCESSION T35997  
VERSION T35997.1 GI:618095  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 253)  
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,  
Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
FitzGerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geodhagen,N.S.M.,  
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle Jr,P.S.,  
Kelley,J.M., Kline,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,  
Merrick,J.M., Moreno-Falange,R.F., McDonald,L.A., Nguyen,D.T.,  
Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,  
Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,  
Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A.,  
Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrie,A.,  
Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,  
Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,  
Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C.,  
Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,  
Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.  
Initial Assessment of Human Gene Diversity and Expression Patterns  
Based Upon 83 Million Basepairs of cDNA Sequence  
Nature 377, 3-174 (1995)  
96026280  
Other ESTs: THC10234  
Contact: Venter, JC  
The Institute for Genomic Research  
932 Clopper Rd, Gaithersburg, MD 20878  
Tel: 3018699056  
Fax: 3018699423  
Email: tdbinfo@tdb.tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please contact the TIGR Database  
(tdbinfo@tdb.tigr.org)

```

/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: Brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI; sex:Female; dev_stage=3 months old;
Isolate:muscular atrophy patient; tissue_type=total
brain; total mRNA was oligo-(dT) primed and directionally
cloned: 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Soares, Psychiatry
Dept. Columbia University, USA.
Bento Soares, P.N.A.S in press"
61 a 58 c 87 t 2 others

```

[illegible]

RESULT	8	
F05255		
LOCUS	272 bp	18-FEB-1995
DEFINITION	HSC05B041 normalized infant brain cDNA Homo sapiens CDNA clone	
	C-05B04, mRNA sequence.	
ACCESSION	F05255	
VERSION	F05255.1	
KEYWORDS	GI:668503	
SOURCE	EST.	
ORGANISM	human.	
	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
	Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
REFERENCE	1 (bases 1 to 272)	
AUTHORS	Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C.,	
	Devignes, M.D., Duprat, S., Houlgatte, R., Jumeau, M.N., Lamy, B.,	
	Lorenzo, F., Mitchell, H., Marilgate-Samson, R., Pietu, G., Pouliot, Y.,	
	Sebastiani-Kabaktchis, C. and Tessier, A.	
TITLE	IMAGE: molecular integration of the analysis of the human genome	
	and its expression	
	sci. data. 117	263-272 (1995)

JOURNAL MEDLINE COMMENT

and its expression  
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
95277534  
Genexpress-Genethon  
Contact: Genethon  
Genethon Centre de recherche sur le Genome Humain  
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
Tel: 33169472800  
Fax: 33160778698  
Email: genexpress@genethon.fr  
Single read.  
Genexpress\_library\_idt: C; Genexpress\_sequence\_idt: ylc-05B04  
Insert Length: 2281 Std Error: 0.00  
Seq primer: (-21)M13\_universal  
High quality sequence stop: 233.  
Location/Qualifiers  
1. .272  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="c-05B04"  
/clone\_lib="normalized infant brain cdna"  
/sex="Female"

FEATURES source

RESULT	9
T32119	
LOCUS	EST43807 Human Brain Homo sapiens cDNA 5' end similar to None, mRNA
DEFINITION	t32119 330 bp mRNA EST 06-SEP-1995
ACCESSION	T32119
VERSION	T32119.1 GI:614217
KEYWORDS	sequence.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
	Carnivora; Primates; Hominidae; Homo.

REFERENCE  
AUTHORS

1. (bases 1 to 330)  
Euterria; Primates

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
Bult,C., Lee,N., Kirkness,E.F., Weinstein,K.G., Gocayne,J.D.,  
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,  
Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
FitzGerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,  
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle Jr,P.S.,  
Kelley,J.M., Klinek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,  
Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T.,  
Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,  
Saudke,D.N., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,  
Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A.,  
Coleman,T.A., Collins,E.J., Dimke,D., Feng,P., Ferrle,A.,  
Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,  
Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,  
Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wang,J., Xu,C.,  
Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,  
Haseitine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.

Initial Assessment of Human Gene Diversity and Expression Patterns  
Based Upon 83 Million Basepairs of cDNA Sequence  
Nature 377 3-174 (1995)

TITLE

JOURNAL

96026280  
MEDLINE  
Other\_ESTs: THC10234  
Contact: Venter, JC  
The Institute for Genomic Research  
932 Clionner Rd. Gaithersburg MD 20878



```

/db_xref="GDB:5276577"
/db_xref="taxon:9606"
/clone="IMAGE:649988"
/clone_lib="stratagene NT2 neuronal precursor 937230"
/tissue_type="neuroepithelial cells"
/dev_stage="Ntera-2 neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: brain; Vector: pBluescript SK-; Site_1:
ECORI; Site 2: XhoI; Cloned unidirectionally. Primer:

```

[illegible]

```

RESOLUTION 11
AA2233940
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

AA2233940 485 bp EST
zr13f02.r1 StrataGene hNT neuron 11-MAR-1998
IMAGE:648699 5', mRNA sequence.

AA2233940.1 GI:1844569
EST.
human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 485)
Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)

On Nov 29, 1993 this sequence version replaced gi:430348.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watonst.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2104 Std Error: 0.00
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 403.
Location/Qualifiers
1. .485
/organism="Homo sapiens"
/db_xref="GDB:5589219"

```

```

/db_xref="taxon:9606"
/clone="IMAGE:648699"
/clone_lib="Stratagene hNT neuron (#937233)"
/dev_stage="hNT neurons"
/lab_host="SOLR (kanamycin resistant)"
/notes="vector: pluescript SK; site_1: EcoRI; site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT.
Differentiated, post mitotic hNT neurons. Average insert
size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTT 3'"
BASE COUNT      137 a   92 c   117 g   137 t   2 others
ORIGIN

Query Match      23.8%; Score 203; DB 30; Length 485;
Best Local Similarity 100.0%; Pred. No. 7.7e-84;
Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaggatggggagcagctgaatccgagatggatgaaccgtttgtcacagcattgtga 60
Db 14 GGAGGATGGCGAGCAGCTGTAATCCAGATGGATGGAATCCGTTTGTCTACAGCATTTGTA 73

QY 61 attgttgatgtctagctcatctccaccatctacatggcagcctccattggcacagac 120
Db 74 ATTGCTGTGTGCTGTAGCCTCATTTCCACCATCTACATGGCAGCCTCCATTGGCACAGAC 133

QY 121 ttctggtatgaatcgaagtcaggtccaagaaatccagtgatttgtaaaaaagcatc 180
Db 134 TTCTGATGATGATTCGAAGTCCAGTTCAGAAATAATCCAGTGATTGTAATAAAGCATC 193

QY 181 tggagtgaaattcattagtagtga 203
Db 194 TGGGATGAATTCATTAGTCATGA 216

RESULT 12
D45319      250 bp      mRNA      EST      30-DEC-1995
LOCUS      HUMHG5097 Human cerebral cortex Homo sapiens cDNA, mRNA sequence.
DEFINITION D45319
ACCESSION  D45319
VERSION    D45319.1 GI:1136721
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 250)
AUTHORS   Takahashi, N., Hashida, H., Zhao, N., Misumi, Y. and Sakaki, Y.
TITLE     High-density cDNA filter analysis of the expression profiles of the
          genes preferentially expressed in human brain
JOURNAL   Gene 164, 219-227 (1995)
MEDLINE   96089586
COMMENT   Contact: Nobuaki Takahashi
          Institute of Medical Science
          University of Tokyo
          Shirokanedai 4-6-1, Minato-ku, Tokyo, Japan 108
          Tel: 03-5449-5625
          Fax: 03-5449-5445
          Insert length: 563 Std Error: 0.00.
          Location/Qualifiers
            1..250
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone_lib="Human cerebral cortex"
              /note="Adult male cerebral cortex tissue."
          /a 69 a 47 c 60 g 72 t 2 others

BASE COUNT      69 a   47 c   60 g   72 t   2 others
ORIGIN

Query Match      23.7%; Score 202; DB 24; Length 250;
Best Local Similarity 100.0%; Pred. No. 2.6e-83;
Matches 202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

## ORIGIN

Query Match 23.4%; Score 200; DB 27; Length 456;  
Best Local Similarity 100.0%; Pred. No. 1.9e-82;  
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ggatggcgagcagctgaatgccagaatggataaccgtttgtgtacagcatttgaatt 63  
|||||  
Db 1 GGATGGCGAGCAGCTGTAATGCCAGATGGATAACCGTTTGTGTACAGCATTGTGAATT 60  
|||||

QY 64 gcttggtgcttagcctcatttcaccattctacatggcagcctccatggcacagacttc 123  
|||||  
Db 61 GCTTGTGCTTAGCCTCAATTTCCACCATCTACATGGCAGCCTCCATTTGGCAGACTTC 120  
|||||

QY 124 tggatgaatccgaagtcagctccagttcaagaaattccagtgatttgaataaaagcatctgg 183  
|||||  
Db 121 TGGTATGAATATCGAAGTCAGTTCAGAAATTCAGTGATTGTAATAAAGCATCTGG 180  
|||||

QY 184 gatgaattcattagtgatga 203  
|||||  
Db 181 GATGAATTCATTAGTGATGA 200  
|||||

## RESULT 14

AA040832 AA040832 514 bp mRNA EST 30-AUG-1996  
LOCUS zk47f04.r1 Soares\_pregnant\_uterus\_NbHPU Homo sapiens cDNA clone  
DEFINITION IMAGE:485983 5', mRNA sequence.

ACCESSION AA040832  
VERSION AA040832.1 GI:1517128  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.

## REFERENCE

1 (bases 1 to 514)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.

## TITLE

The WashU-Merck EST Project

## JOURNAL

Unpublished (1995)

## COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 396.

## FEATURES

## source

1. 514  
/organism="Homo sapiens"  
/db\_xref="GDB:3759697"  
/db\_xref="taxon:9606"  
/clone="IMAGE:485983"  
/clone\_lib="Soares\_pregnant\_uterus\_NbHPU"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"

/note="Organ: uterus; Vector: pT73-Pac; Site\_1: Not I;  
Site\_2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo(dT) primer [5,  
AATCGAAGATTCGGCGCCGCTTTTTTTTTTTTTTTT 3'];  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by M. Fatima Bonaldo."

BASE COUNT  
ORIGIN

137 a 97 c 124 g 154 t 2 others

Query Match 23.4%; Score 200; DB 27; Length 514;  
Best Local Similarity 100.0%; Pred. No. 1.9e-82;  
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ggatggcgagcagctgaatgccagaatggataaccgtttgtgtacagcatttgaatt 63  
|||||  
Db 1 GGATGGCGAGCAGCTGTAATGCCAGATGGATAACCGTTTGTGTACAGCATTGTGAATT 60  
|||||

QY 64 gcttggtgcttagcctcatttcaccattctacatggcagcctccatggcacagacttc 123  
|||||  
Db 61 GCTTGTGCTTAGCCTCAATTTCCACCATCTACATGGCAGCCTCCATTTGGCAGACTTC 120  
|||||

QY 124 tggatgaatccgaagtcagctccagttcaagaaattccagtgatttgaataaaagcatctgg 183  
|||||  
Db 121 TGGTATGAATATCGAAGTCAGTTCAGAAATTCAGTGATTGTAATAAAGCATCTGG 180  
|||||

QY 184 gatgaattcattagtgatga 203  
|||||  
Db 181 GATGAATTCATTAGTGATGA 200  
|||||

## RESULT 15

AA224086 AA224086 602 bp mRNA EST 11-MAR-1998  
LOCUS zrl3g03.r1 Stratagene hnt neuron (#937233) Homo sapiens cDNA clone  
DEFINITION IMAGE:648724 5', mRNA sequence.

ACCESSION AA224086  
VERSION AA224086.1 GI:1844637  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.

## REFERENCE

1 (bases 1 to 602)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.

## TITLE

WashU-NCI human EST Project

## JOURNAL

Unpublished (1997)

## COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 2097 Std Error: 0.00  
Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 340.

## FEATURES

## source

1. 602  
/organism="Homo sapiens"  
/db\_xref="GDB:5589244"  
/db\_xref="taxon:9606"  
/clone="IMAGE:648724"  
/clone\_lib="Stratagene hnt neuron (#937233)"  
/dev\_stage="hnt neurons"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Cloned unidirectionally. Primer: Oligo dt.  
Differentiated, post mitotic hnt neurons. Average insert  
size: 1.5 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5'  
GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'  
CTCAGCTTTTTTTTTTTTTTTT 3"

## BASE COUNT

## ORIGIN

153 a 129 c 146 g 171 t 3 others

Fri May 12 12:27:30 2000

Query Match 23.4%; Score 200; DB 30; Length 602;  
Best Local Similarity 100.0%; Pred. No. 1.8e-82;  
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	4	ggatggcgagcagctctgaatgccagaatggataaccgcttttgctacagcatttgaatt	63
Db	18	GGATGGCGAGCAGCTCTGAATGCCAGATGGATAACCGTTTGTCTACAGCATTTGTAAT	77
QY	64	gcttgctgcttagcctcattccaccatctacatggcagcctccattggcacagacttc	123
Db	78	GCTTGCTGCTTAGCCTCATTTCCACCACTACATGGCAGCCTCCATTGGCACAGACTTC	137
QY	124	tggatgaatatcgaaagtcagttccagaaattccagtgatttgaataaaagcactcgg	183
Db	138	TGGTATGAATATCGAAGTCCAGTTCAAGNAAATTCAGTGATTGTAATAAAGCATCTGG	197
QY	184	gatgaattcattagtatga	203
Db	198	GATGAATTCATTAGTGATGA	217

Search completed: May 1, 2000, 14:44:56  
Job time: 14016 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:53:11 ; Search time 4425.31 Seconds  
(without alignments)  
-566.752 Million cell updates/sec

Title: US-09-215-435-124

Perfect score: 826

Sequence: 1 aaccagaggtgccatgggt.....atcatccaaaaaaaaaaaaa 826

Scoring table: OLIGO\_NVC  
Gapop 60.0 , Gapext 60.0

Searched: 821193 seqs, -1518192014 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba1.\*

2: gb\_ba2.\*

3: gb\_om.\*

4: gb\_ov.\*

5: gb\_pat.\*

6: gb\_ph.\*

7: gb\_p11.\*

8: gb\_p12.\*

9: gb\_p13.\*

10: gb\_p14.\*

11: gb\_p15.\*

12: gb\_ro.\*

13: gb\_sts.\*

14: gb\_sy.\*

15: gb\_un.\*

16: gb\_v1.\*

17: em\_fun.\*

18: em\_hum1.\*

19: em\_hum2.\*

20: em\_in.\*

21: em\_om.\*

22: em\_or.\*

23: em\_ov.\*

24: em\_pat.\*

25: em\_ph.\*

26: em\_pl.\*

27: em\_ro.\*

28: em\_sts.\*

29: em\_sy.\*

30: em\_un.\*

31: em\_v1.\*

32: gb\_htg1.\*

33: gb\_htg2.\*

34: gb\_in1.\*

35: gb\_in2.\*

36: em\_ba1.\*

37: em\_ba2.\*

38: em\_hum3.\*

39: em\_hum4.\*

40: gb\_pr4.\*

41: gb\_htg3.\*

42: gb\_htg4.\*

43: gb\_htg5.\*

44: gb\_htg6.\*  
45: gb\_htg7.\*  
46: em\_htg1.\*  
47: em\_htg2.\*  
48: em\_htg3.\*  
49: em\_hum5.\*  
50: gb\_pl3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	208	25.2	352	13	G27363	G27363 human STS S
C 2	23	2.8	105615	41	AC011345	AC011345 Homo sapi
C 3	23	2.8	142203	41	AC008726	AC008726 Homo sapi
C 4	23	2.8	149554	10	HSAC000111	AC000111 Human BAC
C 5	22	2.7	14968	34	CELW03A3	U50184 Caenorhabdi
C 6	21	2.5	13951	35	AE001406	AE001406 Plasmodi
C 7	21	2.5	24942	34	CEK03D3	282276 Caenorhabdi
C 8	21	2.5	60366	43	AC013768	AC013768 Homo sapi
C 9	21	2.5	110000	32	CEY113B8_1	Continuation (2 of
C 10	21	2.5	110000	32	CEY113B8_2	Continuation (3 of
C 11	21	2.5	163678	41	AC010178	AC010178 Homo sapi
C 12	21	2.5	166214	33	AC006735	AC006735 Caenorhab
C 13	21	2.5	176586	44	AC009306	AC009306 Homo sapi
C 14	21	2.5	185000	43	AC007799	AC007799 Homo sapi
C 15	21	2.5	224129	32	CEY7A9	X95312 Caenorhabdi
C 16	20	2.4	1152	4	ONTGFB	X99303 O.mykiss mr
C 17	20	2.4	1466	34	TCU41444	U41444 Trypanosoma
C 18	20	2.4	8699	34	PFSC03018	AL008971 Plasmodi
C 19	20	2.4	18212	7	AP000388	AP000388 Arabidops
C 20	20	2.4	30020	35	CEL07H3	AF077540 Caenorhab
C 21	20	2.4	39519	42	AC012054	AC012054 Homo sapi
C 22	20	2.4	66531	43	AC015829	AC015829 Homo sapi
C 23	20	2.4	85626	20	AC005113	AC005113 Drosophil
C 24	20	2.4	86829	34	PFMAL3P5	AL034556 Plasmodi
C 25	20	2.4	95824	8	ATU90439	U90439 Arabidops
C 26	20	2.4	113714	43	AC015766	AC015766 Homo sapi
C 27	20	2.4	151299	43	AC012592	AC012592 Homo sapi
C 28	20	2.4	156045	43	AC009468	AC009468 Homo sapi
C 29	20	2.4	158400	42	AC011029	AC011029 Homo sapi
C 30	20	2.4	159179	12	MMHG322F16	AF111103 Mouse maj
C 31	20	2.4	159957	10	AP000475	AP000475 Homo sapi
C 32	20	2.4	161754	35	AC007451	AC007451 Drosophil
C 33	20	2.4	174916	42	AC012102	AC012102 Homo sapi
C 34	20	2.4	179305	44	AC013532	AC013532 Homo sapi
C 35	20	2.4	190959	46	AC006246	AC006246 Drosophil
C 36	20	2.4	207277	43	AC015679	AC015679 Homo sapi
C 37	20	2.4	245805	33	AC006752	AC006752 Caenorhab
C 38	20	2.4	247774	33	AC004803	AC004803 Homo sapi
C 39	20	2.4	252669	42	AC007782	AC007782 Homo sapi
C 40	20	2.4	278007	33	AC006799	AC006799 Caenorhab
C 41	19	2.3	461	3	OA2204T	X84291 O.aries gen
C 42	19	2.3	629	10	IROEST026	AL110289 Homo sapi
C 43	19	2.3	1053	10	HSM800234	AL049449 Homo sapi
C 44	19	2.3	1220	40	HSSPGSC14	AF080524 Homo sapi
C 45	19	2.3	1303	34	ECCU1SNRNA	M73768 E.mutillocu

# ALIGNMENTS

RESULT 1  
G27363/c G27363 352 bp DNA STS  
LOCUS human STS SHGC-31033, sequence tagged site.  
DEFINITION G27363  
ACCESSION G27363  
VERSION G27363.1 GI:1396086

28-JUN-1996

KEYWORDS

ST5; ST5 sequence; primer; sequence tagged site.

SOURCE

human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 352)

AUTHORS

Myers,R.M.

JOURNAL

Unpublished (1996)

COMMENT

Contact: Richard M. Myers  
Stanford Human Genome Center (SHGC)  
Stanford University School of Medicine  
Department of Genetics, M-344, Stanford, CA 94305, USA  
Tel: 4157259687  
Fax: 4157259689  
Email: myers@shgc.stanford.edu  
  
Primer A: TTGGAAGAAGAGGGGTTCTG  
Primer B: AAGCACAAAACACGAGCG  
ST5 size: 133  
PCR Profile:  
Initial incubation: 94 degrees C for 90 seconds  
Denaturation: 94 degrees C for 15 seconds  
Annealing: 62 degrees C for 23 seconds  
Polymerization: 72 degrees C for 30 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9600  
  
Protocol:  
Template: 25 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
Tag Polymerase: 0.05 units/uL  
Total Vol: 10 uL  
  
Buffer:  
MgCl2: 2.5 mM  
KCl: 50 mM  
Tris-HCl: 20 mM  
pH: 8.3  
  
Prepared with primer pairs provided by Sandoz, derived from H83927  
-- Washington University/Merck EST sequence.  
  
FEATURES  
source  
1..352  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="8"  
4..136  
primer\_bind  
4..133  
primer\_bind  
complement(119..136)  
BASE COUNT  
69 a 76 c 110 g 96 t 1 others  
ORIGIN  
25.2%; Score 208; DB 13; Length 352;  
Query Match  
Best Local Similarity 100.0%; Pred. No. 3.5e-110;  
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 495 gtcatctctctcccaaggaaacaaactcgaggtctcttgaaatggacagattt 554  
|||||  
Db 304 GTCACTCTCTCTCCCAAGAAACAAACTCGAGGCTCTTGAAATGGACAGATT 245  
|||||  
Qy 555 ctgaacggtttccacctggggaacctgaagcagccagttcatgaccagaactac 614  
|||||  
Db 244 CTGAACCGTTTCCACCTGGGCGAACCCTGAAGCAAGCACCAGCTTCATGACCCAGACTAC 185  
|||||  
Qy 615 caggactcaccaaccttcaggtctccagaaagagccagcagcccaagcacaacac 674  
|||||  
Db 184 CAGGACTCACCAACCGCTCAGGCTCCCAAGAAAGGGCCGAGCCCAACCAACAAAC 125  
|||||  
Qy 675 caggcgagatagctgctctgtagatag 702  
|||||

Db 124 CAGCGGAGATAGTGCCTCTAGATAG 97

RESULT 2

AC011345/c

LOCUS

AC011345.1

DEFINITION

Homo sapiens chromosome 5 clone C17-HSPC-293E5, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 7 unordered pieces.

ACCESSION

AC011345

VERSION

AC011345.1

KEYWORDS

HTG; HTGS\_PHASE1.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 105615)

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

JOURNAL

DOE Joint Genome Institute.

REFERENCE

2 (bases 1 to 105615)

AUTHORS

DOE Joint Genome Institute.

JOURNAL

Direct Submission

COMMENT

Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
www.jgi.doe.gov.  
\* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
1  
643: contig of 643 bp in length  
gap of unknown length  
644 1942: contig of 1299 bp in length  
gap of unknown length  
1943 4619: contig of 2677 bp in length  
gap of unknown length  
4620 20310: contig of 15691 bp in length  
gap of unknown length  
20311 37377: contig of 17067 bp in length  
gap of unknown length  
37378 69427: contig of 32050 bp in length  
gap of unknown length  
69428 105615: contig of 36188 bp in length.  
Location/Qualifiers  
1..105615  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="C17-HSPC-293E5"  
BASE COUNT  
30560 a 20772 c 20258 g 34004 t 21 others  
ORIGIN

Query Match  
2.8%; Score 23; DB 41; Length 105615;  
Best Local Similarity 100.0%; Pred. No. 0.059;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 30 aggtggtcacagcagcactgtt 52  
|||||  
Db 95597 AGGCTGTCACAGCAGCAGTGT 95575  
|||||  
  
RESULT 3  
AC008726

LOCUS

AC008726

DEFINITION

Homo sapiens chromosome 5 clone C17978SKB\_98D10, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 18 unordered pieces.

ACCESSION

AC008726

VERSION

AC008726.1

KEYWORDS

HTG; HTGS\_PHASE1.

SOURCE ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 142203)

TITLE DOE Joint Genome Institute.

JOURNAL Sequencing of Human Chromosome 5

REFERENCE 2 (bases 1 to 142203)

AUTHORS Unpublished

TITLE DOE Joint Genome Institute.

JOURNAL Direct Submission

COMMENT Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA www.jgi.doe.gov.

\* NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 671: contig of 671 bp in length

672 gap of unknown length

1724: contig of 1053 bp in length

1725 gap of unknown length

3020: contig of 1296 bp in length

3021 gap of unknown length

4019: contig of 999 bp in length

4020 gap of unknown length

6632: contig of 2613 bp in length

6633 gap of unknown length

9136: contig of 2504 bp in length

9137 gap of unknown length

12586: contig of 3450 bp in length

12587 gap of unknown length

16174: contig of 3588 bp in length

16175 gap of unknown length

27128: contig of 10954 bp in length

27129 gap of unknown length

34733: contig of 7605 bp in length

34734 gap of unknown length

44792: contig of 10059 bp in length

44793 gap of unknown length

55589: contig of 10797 bp in length

55590 gap of unknown length

67363: contig of 11774 bp in length

67364 gap of unknown length

79085: contig of 11722 bp in length

79086 gap of unknown length

92320: contig of 13235 bp in length

92321 gap of unknown length

107751: contig of 15431 bp in length

107752 gap of unknown length

124346: contig of 16595 bp in length

124347 gap of unknown length

142203: contig of 17857 bp in length.

Location/Qualifiers

1. 142203

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

/clone="C19785KB\_98D10"

BASE COUNT 42895 a 27899 c 27778 g 43580 t 51 others

ORIGIN

Query Match 2.8%; Score 23; DB 41; Length 142203;

Best Local Similarity 100.0%; Pred. No. 0.06;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 agcctggtcacagcagcactgtt 52

|||||

Db 82611 AGGCTGTCACAGCAGCAGCTCTT 82633

RESULT 4

LOCUS HSAC000111 149554 bp DNA PRI 31-JAN-1997

DEFINITION Human BAC clone 068P20 from 7q31-q32, complete sequence.

ACCESSION AC000111

VERSION AG000111.1 GI:1809237

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 149554)

TITLE Scheet, P., Maggi, L. and Blandford, M.

JOURNAL The sequence of H. sapiens BAC clone 068P20

REFERENCE Unpublished (1997)

AUTHORS 2 (bases 1 to 149554)

TITLE Waterston, R.

JOURNAL Direct Submission

COMMENT Submitted (31-JAN-1997)

Genome Sequencing Center

Department of Genetics, Washington University

St. Louis, MO 63108, USA

http://genome.wustl.edu/gsc

e-mail: sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).

VECTOR: pBelo

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right of H\_068P20 is H\_133K23, 200 bp overlap. Actual start of this clone is at base position 1 of HUM068P20; actual end is at 33944 of HUM133K23. This clone is part of an unanchored island, orientation is unknown.

This clone contains STS SW55115 (NID:G394430).

Location/Qualifiers

1. 149554

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="7"

/clone="068P20"

/map="7q31-q32"

/clone\_lib="CITB-978SK-B"

/map="7q31-q32"

complement(360..825)

/rpt\_family="L1"

complement(1587..2370)

/rpt\_family="L1"

2362..3556

/rpt\_family="L1"

complement(2944..3337)

/rpt\_family="L1"

3236..3508

/note="GRAIL prediction, score = 84"

repeat\_region

repeat\_region

repeat\_region

repeat\_region

exon

FEATURES

source

```
/evidence-not_experimental
4669. .4710
/rpt_family="L1"
4798. .5090
/rpt_family="ALU"
5716. .5997
/rpt_family="ALU"
5998. .6062
/rpt_family="L1"
6526. .6778
/rpt_family="ALU"
complement(7199. .7291)
/rpt_family="L1"
7863. .7883
/rpt_family="L1"
complement(7882. .8025)
/rpt_family="L1"
complement(8840. .9368)
/rpt_family="L1"
complement(9375. .9645)
/rpt_family="ALU"
10079. .10098
/rpt_family="L1"
complement(10988. .11280)
/rpt_family="ALU"
complement(14250. .14272)
/rpt_family="L1"
14929. .15003
/!note="77% similarity to Human clone 282291
(NID:g1666400)"
15025. .15219
/!note="match to Human cystic fibrosis gene, region
upstream of the promoter U48778 (NID:g1215738)"
complement(15220. .15512)
/rpt_family="ALU"
complement(15513. .15664)
/rpt_family="MIR"
16734. .16993
/rpt_family="MIR"
19177. .19305
/rpt_family="PTR"
complement(19251. .19343)
/rpt_family="PTR"
complement(19776. .19829)
/rpt_family="PTR"
19802. .>146461
/!gene="cfr"
join(19802. .19854, 43960. .44070, 48741. .48849, 70606. .70821,
73983. .74072, 74955. .75118, 76255. .76380, 79807. .80053,
81723. .81815, 88348. .88530, 99171. .99362, 127446. .127540,
130060. .130146, 131641. .132364, 134637. .134765,
142533. .142570, 143239. .143489, 146381. .>146461)
/!gene="cfr"
/!note="match to M5131 (NID:g306536), this gene continues
into H_133K23 at base 150226"
/codon_start=1
/product="transmembrane chloride conductor protein"
/db_protein_id="AAB46352.1"
/db_xref="GI:1809238"
/translation="WQSPLEKASVYKLFSSWTRPILRKGYRQLRSLDIYQIPSD
SADNLSKLEREDWELASKNPKLNALRCFFWFEMFYGIFLVIGETKAVQPLL
GRILASYPDNKEERSIAYLIGICLLFIVRTLLHPAIFGLHIGQMRIAMESLI
YKTKLCSRVLDKISIGQLSVLSNNKRFDEGLAHFWIAPLOVALLMGLIWEI
LQASAFGLGFLIVIALFOAGLNNMKYDQDQAGKISERUITSEMENIQSVKAYC
WEAKMEKTEIQRQELKTRKAAVYFNSSAFFSGFVFLSVLYALIKGILR
KIFTTISFCIVLRMAVTRQFPWAVQWYDSLGAINKIQDFLOKQYKTYLEYNLTTEV
VVENTAVWEEGEGFELFEKAKNNNRKTSNGDLSLFFSNFLGTPVVKDINFKIER
GOLLAVAGSTGAGKSLAMVINGELEPSEGGIKHSGRISFCFSQFSLMPGTIKENIF
GYSYDEYRYSVIKACQLEDISKFAEKDNIVLGGGIIISGQQRARISLARAYKDA
DYLDSLPQDLVTEKEIFESCCKLMANKTRILVTSKMEHLKADKILLIHGESS
YVGTFSQLQIDPDSKLMGCDSPQFSAERNISILFETLHRFSLEGDAPVSTET
KQSFQKTFEGEGRKNSILNINSIKFSIVQKTPLOMNGIEEDSDEPLRSLVLP
DSEQGAAILPRISVISTGPTLOARRRQSVLNLMTSHVNGQNIHRKTTASTRKSVLAP
```

```
QANTELDIYRRLSQETGLEISEEINEDLKECFFDDMESIPATVWNTYLRIVTH
KSLIFVLWGLVIFLAEEAASLVVLLGNTPLQDKGNTSHSRNSXAVITSSYY
VFYIVGVADTLLAMGFRGLPLVTLITVSKILHKKMLHSLVQAPMSTLTLKAGGI
LNRFSKDIAILDLLPLTFIDFIQ"
20137. .20159
/rpt_family="L1"
complement(23276. .23580)
/rpt_family="ALU"
complement(25398. .25417)
/rpt_family="L1"
27466. .27639
/!gene="cfr"
/!note="possible L32E Ribosomal Protein psuedogene"
27504. .27525
/rpt_family="L1"
complement(29718. .29762)
/!gene="cfr"
/!note="GRAIL prediction, score = 86"
/evidence-not_experimental
30664. .30960
/rpt_family="MER"
complement(31099. .31561)
/rpt_family="THE"
35372. .35400
/rpt_family="L1"
complement(35569. .35918)
/rpt_family="L1"
complement(35919. .36209)
/rpt_family="ALU"
complement(36213. .36604)
/rpt_family="L1"
complement(36675. .36707)
/rpt_family="L1"
complement(37143. .37250)
/!gene="cfr"
/!note="GRAIL prediction, score = 90"
/evidence-not_experimental
complement(38497. .38567)
/rpt_family="L1"
complement(38749. .38796)
/rpt_family="L1"
39378. .39548
/rpt_family="ALU"
complement(39973. .39999)
/rpt_family="L1"
complement(40001. .40296)
/rpt_family="ALU"
41171. .41464
/rpt_family="L1"
41467. .41529
/rpt_family="L1"
complement(43675. .43704)
/rpt_family="L1"
44180. .44474
/rpt_family="ALU"
complement(45962. .46251)
/rpt_family="ALU"
complement(46507. .46604)
/rpt_family="L1"
Query Match 2.8% Score 23; DB 10; Length 149554;
Best Local Similarity 100.08; Pred.No. 0.061;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 788 ttctccaaataaaaaaatca 810
|||||
Db 62534 TTTTCCAAATAAAAAAATCA 62556
|||||
RESULT 5
CELW03A3 14968 bp DNA INV 30-DEC-1997
LOCUS Caenorhabditis elegans cosmid w03A3.
DEFINITION
```



ACCESSION	U50184	/chromosome="III"	gene
VERSION	U50184.1	/clone="W03A3"	
KEYWORDS		/complement(3668..10292)	
SOURCE		/gene="W03A3.2"	CDS
ORGANISM	Caenorhabditis elegans strain-Bristol N2.	complement(join(3668..3849,4228..4437,4489..4822,4944..5612,5668..5825,5966..6043,6090..6318,6365..6468,6695..7452,8474..8578,9009..9147,9281..9350,9541..9774,9829..9906,10014..10292))	
REFERENCE	1 (bases 1 to 14968)	/gene="W03A3.2"	
AUTHORS	Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connel,M., Copsey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Eaveil,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lighning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Showkeen,R., Smaldon,N., Smith,A., Sonhammer,E., Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Sproat,J. and Wohlman,P.	/note="Similar to DNA polymerase; coded for by C. elegans cDNA yk47d11.5; coded for by C. elegans cDNA yk47d11.3; coded for by C. elegans cDNA yk130c1.3; coded for by C. elegans cDNA yk134d9.3; coded for by C. elegans cDNA yk47d11.5; coded for by C. elegans cDNA yk134d9.5"	
TITLE	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans	/codon_start=1	
JOURNAL	Nature 368 (6466), 32-38 (1994)	/protein_id="AAB93325.1"	
MEDLINE	94150718	/db_xref="GI:2731173"	
REFERENCE	2 (bases 1 to 14968)	/translation="MLRNSFISQDENDDQSLPTOLGRAIASLSPASLAIFEDLN SAGRALADTELHMLYLVFYKNSRAQIKIKFIYSIFILKFKKLEPKFKKISEN ITVHTNSIRKKQHFVHTVPINVSVMQECMDHHLFISFKLPSDKHRIAKLVGVSEK FILDQQRNDKLLQIHRFFSALADLISENSIYVSKYIPRGCLQTQASQAT YAAIMVAFCLRLGTYLKALDGFATRLFGVSELSVVAIEGIDQARILHERGV TCLSHLSACDSSKLAHFLTLAVPYSSNSNDGLGEWLEGPVRMRVDAARTLREARK VLIRVOELGISVELPKFENEENIQESCDSGLPDCGEMEDELKENVKEWETK SVTEMLTNTISFKSDDLFKKEIYVEDEVFKEIDEDEEVEETVIECLELSL LKLSATDEVFLRRLSQFSGRSILNLSLEDSFDPVPRSSPIILNFIPTPKRE SPTTFYDFSDRPIPGSLPISSSRKSVLTNIANLDSRRSINSNASDNNSDVFT PPTKSAKEKRRIAVKPRVGNIIYSPLTSSPVIKHPKLEINHFIKLDVCHDNAML WTKSSTSTSCSIRSDVDTGIAITDAGNTFIPLETFGGNSVQPFKNINFYVL KMPLKIFWQPSKASYEFSEKCIIPLTNLEFLKTLAVTVMYISSMEDALIFEK FGKIFRLKAVRIAALYNVIDVEQENSFNFLMDRYSILDPERKTKSSSLHRAA VLEENIWLAYKFNHISNEVANVLYRGLIYPTSGCKPLRHLPTNLKLEQONT QHPVIGKILEYQIHTLQCLMPLAKFIRHCWFEMCTSGRITSLPNLQVPRR ISSDGMAROLFJENSENLIGADYKQLELRVLHNSLNLNLTSDRLFEELS I QWNPDAVKOLCYGLIYMGAKSLSELTEWSTDAEKMLKAFEMPGVRSYINETK EYCKEPTSTIIGRTIIKASGIGERARIERVAVNTIQQSASEIFKFAVDIESK IREFGAQIVLTIHDEVCPEIHVAAASESIENCMQNALHLLRVPVRVSMKTGRSW ADLK"	
JOURNAL	Submitted (38-FEB-1996) Robert Waterston	11462..13438	gene
TITLE	Direct Submission	/gene="ceh-10"	
JOURNAL	Submitted (38-FEB-1996) Robert Waterston	join(11462..11566,11612..11747,11795..12096,12464..12647,13131..13438)	CDS
AUTHORS	Waterston,R.	/gene="ceh-10"	
JOURNAL	Submitted (30-DEC-1997) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	/note="W03A3.1"	
TITLE	Submitted by:	/codon_start=1	
JOURNAL	Genome Sequencing Center	/product="c. elegans homeobox protein ceh-10 (SP:P41935)"	
AUTHORS	Department of Genetics, Washington University, St. Louis, MO 63110, USA, and Sanger Centre, Hinxton Hall	/protein_id="AAB93323.1"	
JOURNAL	Cambridge CB10 IRQ, England	/db_xref="GI:1226249"	
TITLE	e-mail: rw@nemastode.wustl.edu and jes@sanger.ac.uk		
COMMENT	NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.		
FEATURES	This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone		
source	NEIGHBORING COSMID INFORMATION:		
	The 3' cosmid is R12B2, 400 bp overlap. Actual start of this cosmid is at base position 1 of CELW03A3; actual end is at 19903 of CELR12B2		
	NOTES:		
	Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation). Location/Qualifiers		
	1..14968		
	/organism="Caenorhabditis elegans"		
	/strain="Bristol N2"		
	/db_xref="taxon:6239"		
	BASE COUNT		4783 a 2738 c 2498 g 4949 t
	ORIGIN		

```

Query Match      2.7% Score 22; DB 34; Length 14968;
Best Local Similarity 100.0%; Pred. No. 0.2; Mismatches 0; Indels 0; Gaps 0;
Matches 22; Conservative 0;

Qy 783 cctcttttcccaataaaaaa 804
|||||
Db 13767 CcTTCTTTCCAAATAAAAA 13746

RESULT 6
AC001406/c
LOCUS      Plasmodium falciparum chromosome 2, section 43 of 73 of the
DEFINITION complete sequence.
ACCESSION  AE001406 AE001362
VERSION     AE001406.1 GI:3845230
KEYWORDS
SOURCE      malaria parasite P. falciparum.
ORGANISM    Plasmodium falciparum
REFERENCE   1 (bases 1 to 13951)
AUTHORS     Gardner, M.J., Tettein, H., Carucci, D.J., Cummings, L.M., Aravind, L.,
            Koonin, E.V., Shallow, S., Mason, T., Yu, K., Fujii, C., Pederson, J.,
            Shen, K., Jing, J., Aston, C., Lai, Z., Schwartz, D.C., Pertea, M.,
            Salzberg, S., Zhou, L., Sutton, G.G., Clayton, R., White, O.,
            Smith, H.O., Fraser, C.M., Adams, M.D., Venter, J.C. and Hoffman, S.L.
            Chromosome 2 sequence of the human malaria parasite Plasmodium
            falciparum
            Science 282 (5391), 1126-1132 (1998)
JOURNAL     Science 282 (5391), 1126-1132 (1998)
MEDLINE     99021743
REMARK      Erratum: [[published erratum appears in Science 1998 Dec
            4; 282(5395):1827]]
REFERENCE   2 (bases 1 to 13951)
AUTHORS     Gardner, M.J.
DIRECT SUBMISSION
TITLE       Submitted (02-NOV-1998) The Institute for Genomic Research, 9712
            Medical Center Drive, Rockville, MD 20814, USA
FEATURES    Location/Qualifiers
            1. .13951
               /organism="Plasmodium falciparum"
               /db_xref="taxon:5833"
               /chromosome="2"
               complement(1921..8358)
               /gene="PFB0615c"
               complement(join(1921..2310,2496..6749,6961..8358))
               /gene="PFB0615c"
               /note="predicted by GlimmerM"
               /codon_start=1
               /product="predicted membrane associated protein"
               /protein_id="AAC71912.1"
               /db_xref="GI:3845231"
               /translation="WANCYNSLSPHFIFSDNIFCFGLKDGVCIFCNLLNKKLYIIT
            CSQDEYVAQYFVYVVKYKKEEDCKNMTINMLQENKKIKTKTCYIKNVVTN
            KYHFLVYNKYHNYLLCSLSPNSPELLNTFVKTKGKIKSMATKNVFLIK
            KKKILNKKNQMSKILQNVLVSKHTLDSLLMGENDVNVICSKKKKKK
            KTDNKKKGHEIKVDYENKINEKNEKINEEKEEKEEKEEKEEKEE
            NEKINEKEDNKPQYKENDNIPGLDHSQVNIPTFSLNKKPDLKIQFSNI
            ILPKMKIIPDEKIIILLSHKSIYIITKNNDLKNMFIKELIFNSPIITWI
            DNVLFLVNLNELIFLSPKCRNLYFYKINNYSHITSFFYKSRNLYISFTKEIV
            CFKRYETPLAVFKKIOTEGNYIDAKYLFKRPRTINTNHNQSNKDKDNDVIR
            EEDFLRNKNFSQVKKRKNKNYELIFNLLRELKTLNLSNNDYIFEDG
            EINKDLRNKSARSUSVYKYNLNLGLHNNKKRWIKQDIRNNHNNKNCVEDV
            CKNRYEKESIFYEDNNNDNMLSHLYFLKKKKKKPDNPHYNDENVIKLDFVSI
            INLHYLANITSFYMSKYLFDLNGLLYTKRNDGKIYDFLSNCTYCYTKNI
            NKIVDIIINEHDYVMDKKHILKNLSKNNLYINIITEKIQSYNIFSMENCTCIF
            LSLDGSFYINTKHLKYENLQNSNLGNOIYCNFKNNKIQYISVFNKNGYIF
            NGYFYQQIILIFFLIYSHKFFIYLVNTHIVLEKKIHQTHILYKNKQNOQ
            ENLNKKRQSSNLYINFLYIKLNKOYVCLSDSKSVSYFYFFDLPREIKM
            YISDKKKKKKNNNDNKKIYIYRNSKNDNNYKENQNEVENHYDDDDDDGKSIPL
            YTRNIFGCSIGNIYAKIGNIYADYIYNTFYIKDNFNHYMSGSPSSFF
            VSHLEPCVYKMKKKKKQKYTCNMKESESKIDYSTNNHNMQRFFFLKRRKLLK

```

```

NKTEFNDNMKEDKLEEKINEDFVITEGEKKSNKKIKKNTQHNDNNNNNDVFICNSL
YELLNKKSFLLNKHGKGIYINERMHTSELYIDIVTNNILCISFNSVDYPLEI
NPHINRYMPYLNNDIQYIYPLIIRGNNYNNNNMYDLFLIKKKNFLLRNKKEDE
EAILOKEDKDHSTICNPKLIQONODOTVTKVEENVFNTINSNEHISFYLSKWI
EDNTSYIYNDLSLKNMNVFLKIKNDISQNTNRKNFFEDIVCMERKYYLENKWN
NEKMKIYDININMNPHTYINILKNKILLNDVEKTKCIEPQNNHNNINNKLEPKQI
SNMKLNEEKTYILDKNYIHNKNTNYFFDETIIFTFKDNSSQNSILKCKLKIYQ
NRYIYQEKYKKKTKKQKLYIINKLQKCSILNKIDFLSAFKNLYVYFNNQTYG
LIYEEKIKKFKNTKQKLYIINKLQKCSILNKIDFLSAFKNLYVYFNNQTYG
KFCNYISYPSNKSNSHLSNFSYNNLSSNNLSSNNLSSNNLSSNNLSSNNLSSHN
LSHNNLSSNNLSSNNLSSNNLSSNNLSSNNLSSNNLSSNNLSSNNLSSNNL
IKFEFFLHNDKNNLFFLNDYIDYDRLOQDESSLYFYHYPCNINLNFYANLFN
VNPLOGEVDLSDWYLYSPYELFTNSRRRLQYILKILIEQFNDLFEILKKEKNCL
KEINFLINKLSLQDVERKYSYFYSYMMKLFIKNQPMYDKGRYINIPDAIKKDL
LKQLNKNMTEKDKINKENYILQTKYNEQKENINVDSTQYINLKHDMKNTPT"
10112..10760
/gene="PFB0620w"
join(10112..10402,10587..10760)
/gene="PFB0620w"
/note="identified by sequence similarity; putative"
/codon_start=1
/product="YOU2-like small euk. C2C2 zinc finger protein"
/db_xref="GI:3845232"
/translation="MFLSPSPSNHMTNLTNKTNEFSQKIDGMLNNTESLPFOKK
SFMCNCFDNTYNTDFETIGKVNQCQKGFHFVQVQEMQLNQLNSCQSCQCFYK
YSPYAKSNIDGPTIEKEMETCVVYKCFDKEPMLPEISDRHLKLTKEEMK"
11649..13658
/gene="PFB0625w"
11649..13658
/gene="PFB0625w"
/note="predicted by GlimmerM"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC71914.1"
/db_xref="GI:3845233"
/translation="MMNKVCSYLSRISIEFKGYNNMINSNEELIONSVDRCNT
ETCLNKKYKNEECIRKIKIISDNDNMSIFIKGRKLYFNTVYVIEKCDKLN
ILANKYLEGIIIEVQLAGNTFIIPCLNKNILSCFLPQLERGLYHLFFPNKERFI
KLLRPGSDDDIKSPLHVIEITDFSHGLKKNKIDDKKEYIINSTHNNFYTKELI
KLYNNIYNNIYNDYKKNKISLQKNFYLNHNEHFYFNLSKQDFIDHSFT
TKMNSYQHNKDIERKRNQNSLDINNMFISQNLKHKVAQRIPILYKRLLYD
NCTVENKWMFHFTKIFEPFNILSTHFTKSEIKOGYIIFAFNIPITNTNKN
KSVINSYHNEIYKKNKINKISNLSNLSGKEEIGNSYMSFTFLTNSDERNC
VDRLWKYIKTVECNNDISNLYSKNKNVVCPSQPINNKNKFNYSGLKITS
DKVSIFEDDWNEDILPVQKFNVSFEDIDPKKLNELNSVEDINGDILLNDNENDK
NDHVCDPTIKSQNESINGYQYNNNESELIITSMNQNNFYIKDMKKNKKNKKMKKI
SMKYLNNFVSFVLIIDEKIYHSVTPINKFILLFIINWMNFVEK"
BASE COUNT 5303 a 1279 c 1308 g 6061 t
ORIGIN

```

```

Query Match      2.5% Score 21; DB 35; Length 13951;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 798 aaaaaaaatcatccaaaa 818
|||||
Db 570 AAAAAAAATCATCCAAAA 550

RESULT 7
LOCUS      CEK03D3 24942 bp DNA INV 23-NOV-1998
DEFINITION Caenorhabditis elegans cosmid K03D3, complete sequence.
ACCESSION  Z82276
VERSION     Z82276.1 GI:1752956
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 24942)

```

## AUTHORS

Matthews, L.

## TITLE

Direct Submission

## JOURNAL

Submitted (11-NOV-1996) Louis, MO 63110, USA. E-mail:

## REFERENCE

jes@sanger.ac.uk or rwnematode.wustl.edu

## AUTHORS

2 (bases 1 to 24942)

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Showkeen, R., Smaldon, N., Smith, A., Sonhammer, E., Stauden, R., Sulston, J., Therry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Spratt, J., and Wohlman, P.

2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*

## TITLE

Nature 368 (6466), 32-38 (1994)

## JOURNAL

94150718

## MEDLINE

## COMMENT

On Dec 25, 1996 this sequence version replaced gi:1695065. Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.

For a graphical representation of this sequence and its analysis see:-

<http://webcbe.sanger.ac.uk/cgi-bin/display?db=wormaceclass=Sequence&object=K03D3>

Current sequence finishing criteria for the *C. elegans* genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone K03D3 is at 1 in this sequence. The true left end of clone C35D6 is at 24839 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence Z99286. The end of this sequence (24839..24942) overlaps with the start of sequence Z82261.

## FEATURES

## source

Location/Qualifiers

1..24942

/organism="Caenorhabditis elegans"

/db\_xref="taxon:6239"

/chromosome="IV"

/clone="K03D3"

353..1957

/gene="K03D3.10"

join(353..496,541..651,854..945,1385..1570,1753..1957)

/gene="K03D3.10"

/note="similar to Ras family (contains ATP/GTP binding P-loop)"

/codon\_start=1

/protein\_id="CAB05247.1"

/db\_xref="GI:3878217"

/db\_xref="SPTREMBL:O45648"

/translation="MQAIKCVVGDGAVGKTCLLSYTTNAPFGGEYILTVSHFLSGHL RKRNLKLNKLILFYLKNEKLSFKKRLLSVNVFDTYSTNVAVDGRPIN LSLWPTAGQVDQDPRIHSFQTDVFLVCFALNNPASFENVRKWPYEVSHCPNTPILVGTAKDLREDROTIERLRERLPQVSHVTOGYMAKEIKAVKYLECSALTQIGLKQV FDEARTGLTTPQTPQTRAKKSNCTVL"

3207..4569

/gene="K03D3.9"

join(3207..3311,3577..3757,4168..4216,4293..4569)

/gene="K03D3.9"

/codon\_start=1

/protein\_id="CAB05246.1"

gene

/db\_xref="GI:3878216"

/db\_xref="SPTREMBL:O45647"

/translation="MQAIKCVVGDGAVGKTCLLSYTTNAPFGGEYILTVSHFLSGHL RKRNLKLNKLILFYLKNEKLSFKKRLLSVNVFDTYSTNVAVDGRPIN LSLWPTAGQVDQDPRIHSFQTDVFLVCFALNNPASFENVRKWPYEVSHCPNTPILVGTAKDLREDROTIERLRERLPQVSHVTOGYMAKEIKAVKYLECSALTQIGLKQV FDEARTGLTTPQTPQTRAKKSNCTVL"

gene

/gene="K03D3.8"

join(5291..6284,6369..6992,7706..7974)

CDS

/gene="K03D3.8"

/note="similar to Retroviral aspartyl proteases"

/codon\_start=1

/protein\_id="CAB05245.1"

/db\_xref="GI:3878215"

/db\_xref="SPTREMBL:O45646"

/translation="MADKEPPPPAGTITQTSQLLPADPAENLAKAKSAATSQAVSTRAR EQELRVESETEKEILKVCNPLTVDEGRSSHPGLIPEAAQGLSYISIESLDFDTR GKMLFDSGNGMINSIKNYSRLPRVMAILGKSDVKINAFSGTMSDFVSFQSPMDHL GASTEVLTEQKVKSCLLTFLAGTARNIAGDFIKINPDIKDELKMLAEKLLNPVLG FKETQLKQKEEGESVDDFYTRLLKLSAASSTEEVENVKRITLDTFLNGNLQLOE TVRLNLPETTEKAFEFARTVELIKADMLKSKFDPRALAIIGVENRQOQSGESQESK SHGPISSHRYDKPRNMYLASEOQKSLDEIKQEFKFSRRLDRLEVGYSSKTSRVRN AVAAESCVIVPVSTSLGELFTTVKIPQIANGIHAVLDTGAGITLTSKNILQRLGI TKFSPNIDGAVGEGGNKIPLAGSAIINLQVSRILQKYCNDNDTEEDFQVFQGN DLLSRLPMFSDFOYKEVHLGKMDIPYRSEETLSQNIKRVDCCPSIVLTHLQPNLS SIPSDPIKEGSAVERSKILTQSFNLQEPYQAAVVTSHQNRWYSINHSKPSLNK NVRF"

gene

8752..10714

/gene="K03D3.6"

join(8752..10082,10363..10714)

CDS

/gene="K03D3.6"

/note="similar to zinc finger, CCHC class (2 domains)"

/codon\_start=1

/protein\_id="CAB05243.1"

/db\_xref="GI:3878213"

/db\_xref="SPTREMBL:O45644"

/translation="MDPSEIEEPPSPQPERELGMLFDDSQFNPPSVSSPSFSFE TEQASVPGKFIQEIAGHRTQSLDTLPSKVIYRPIVYKGLPEEIRVQPSDSIP YAVTVQTPNTTEKQSHKNSQFQGDCHYCKTKTGMSCKQQAANKRNFRQKSP AVSDMSSGSLSTSTPSSSKATTEQASVTGKIQIPAGHRTLSPDQILSKKEIP YPFTVIRGRODSETEVESSEFSDSTNCSTLSAPTIEKSPKASQSQKQNRLEGVC LYCKLIIGHRYVRKRAHAEVKAFOQGESEVVKLOSTOSRILKLSKLEAKVNLH TSGATKREAEAPLVPVQNPANPAPEPSDDDTATYAFRLKLTQTQSTHVPSPR TDNFKGLCHYKRNAGHMVRDCRKKFAEKRAKGLDQASINNDTQQHPRKRTYKIRL LQLYPTGSAATLRRLTVIITVTRGALLRRLTSAARATFCISFRNLIVSRPST ATCYIRROFPHSRSPFRDLFCLLISLSLCTCLAGHCSLSHLYLNKPIN"

gene

/gene="K03D3.7"

join(10768..10919,11627..11831)

CDS

/gene="K03D3.7"

/note="similar to RAS-RELATED PROTEIN RAC2"

/codon\_start=1

/protein\_id="CAB05244.1"

/db\_xref="GI:3878214"

/db\_xref="SPTREMBL:O45645"

/translation="MRVKTFFYNPAFENVRKWPYEVSHCPNTPILVGTAKDLRE DRDVERLRSLQPSRTPQGYVMAKEIKAVKYLECSALTQIGLKQV FDEAIRTGLTTPQTPQTRAKKSNCTVL"

gene

/gene="K03D3.5"

complement(join(12267..12520,12568..12634))

CDS

/gene="K03D3.5"

/codon\_start=1

/protein\_id="CAB05242.1"

/db\_xref="GI:3878212"

/db\_xref="SPTREMBL:O45643"

/translation="MFQOLFSAVFLVLLLEASTPAAPSRVSVAGLVANGLKKNLAE KIIELEKREYNTIIEKANAGNQKLAQATWNKHQELYHKLFAVTEQKATYKELNKQY HLYRF"

gene

14319..17101

/gene="K03D3.4"

join(14319..14341,14419..14683,14822..14983,16185..16291,16351..16690,16738..16869,16910..17101)

CDS

```
/gene="K03D3.4"
/codon_start=1
/protein_id="CA805241.1"
/db_xref="GI:3878211"
/db_xref="SPTREMBL:O45642"
/translation="MDGSFLVGLNLPNSVKIKFSRVRKFKPCRELKVGGENFKFLE
KNIDGIFKFKFRHFGVKFKIKWIKLIPVYKESFKIWAGKLLKLGCIQKILAIVQHFL
TILFISVLIIIVISVKLSHQASQNSPOQYIMQIIVVFFKFIPIVHLTKT
QGNSSIVYISISILSVLGVWNNIFDMQDHNHSTLNLMAICNLFVIAALLFTA
YGFAPFLVDSLSLIIYFFPAIEGTLTKIQLNIRFLIIYFLCGFMGTRIALT
YKDVLRGLPDSGLVKIRDVYMETFVLNGIFLFSIIINFLGIIEGPKLQKYMFCQ
SPLVFETKISSIYIDDDWDFLEVDLFTPLIIQLSYLITFCNLSVLRLKLSLK
IKGTRVHPETVQ"
complement(18846..20093)
/gene="K03D3.3"
complement(join(18846..19015,19072..19213,19327..19385,
19581..19887,19938..20093))
/gene="K03D3.3"
/codon_start=1
/protein_id="CA805240.1"
/db_xref="GI:3878210"
/db_xref="SPTREMBL:O45641"
/translation="MEFTIQVLVFLSKNHNLCFYMLLISCFAYVILFIFYINVKF
SNGSNDKSLFLIINHFQVITITNIVVILVLSLSSAWVKIKFDSQKORVHCFK
NPKESIVLSLIFASIVFGFSLIIAQLFAHILIELLAIOREIIVFFPKQALTLN
FQKNGFQVSRVQLKNIYIICFTIFLASILHIVFTWKRGTAPHNNLQKXILMOT
ICVGFCCVCVILKLSADIIILDDFFTFPFMIQVCCLMNNRQQLAGLISLFLKMLKFLK
K"
complement(20263..20742)
/gene="K03D3.2"
complement(join(20263..20442,20488..20742))
/gene="K03D3.2"
/codon_start=1
/protein_id="CA805239.1"
/db_xref="GI:3878209"
/db_xref="SPTREMBL:O45640"
/translation="MEFLILFLATISYSOLIAESLVDVVGAGSYDYRTDFTGGEK
TIGIKRLNGLDDEDDKQWLYVECAKAGKANKKCSWVDGKNGKIKGEVTLKILKG
KNVLCQPVVRDAGRVSIFEDSKQTSVQVQVRPAPLHG"
21587..23187
/gene="K03D3.1"
join(21587..21751,21813..21925,21990..22284,22333..22494,
22963..23187)
/gene="K03D3.1"

Query Match          2.5%; Score 21; DB 34; Length 24942;
Best Local Similarity 100.0%; Pred. No. 0.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 797 taaaaaaatcatccaaa 817
|||||
Db 18947 TAAAAAAATCATCCAAA 18967

RESULT 8
AC013768/c
LOCUS          AC013768      60366 bp      DNA          HTG          15-NOV-1999
DEFINITION    Homo sapiens clone 2_K_12, LOW-PASS SEQUENCE SAMPLING.
ACCESSION     AC013768
VERSION       AC013768.1  GI:6425763
KEYWORDS      HTG; HTGS_PHASE0.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 60366)
AUTHORS       Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE         Homo sapiens, clone 2_K_12
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 60366)
AUTHORS       Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
               Baldwin,J., Barna,N., Beckerly,K., Boguslavsky,L., Boukhgaiter,B.,
               Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
```

```
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
* NOTE: This record contains 75 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 794: contig of 794 bp in length
795 1617: contig of 823 bp in length
1618 2432: contig of 815 bp in length
2433 3219: contig of 787 bp in length
3220 4016: contig of 797 bp in length
4017 4800: contig of 784 bp in length
4801 5651: contig of 851 bp in length
5652 6455: contig of 804 bp in length
6456 7236: contig of 781 bp in length
7237 8055: contig of 819 bp in length
8056 8854: contig of 799 bp in length
8855 9653: contig of 798 bp in length
9653 10423: contig of 771 bp in length
10424 11202: contig of 779 bp in length
11203 12007: contig of 805 bp in length
12008 12818: contig of 811 bp in length
12819 13620: contig of 802 bp in length
13621 14427: contig of 807 bp in length
14428 15225: contig of 798 bp in length
15226 16052: contig of 827 bp in length
16053 16857: contig of 805 bp in length
16858 17656: contig of 799 bp in length
17657 18441: contig of 785 bp in length
```

TITLE  
JOURNAL  
COMMENT

```
* * gap of unknown length
* * 19231: contig of 790 bp in length
* * gap of unknown length
* * 20063: contig of 832 bp in length
* * gap of unknown length
* * 20064: contig of 818 bp in length
* * gap of unknown length
* * 20882: contig of 850 bp in length
* * gap of unknown length
* * 21732: contig of 813 bp in length
* * gap of unknown length
* * 22545: contig of 839 bp in length
* * gap of unknown length
* * 23384: contig of 817 bp in length
* * gap of unknown length
* * 24201: contig of 805 bp in length
* * gap of unknown length
* * 25006: contig of 803 bp in length
* * gap of unknown length
* * 25809: contig of 804 bp in length
* * gap of unknown length
* * 26613: contig of 809 bp in length
* * gap of unknown length
* * 27422: contig of 808 bp in length
* * gap of unknown length
* * 28230: contig of 830 bp in length
* * gap of unknown length
* * 29060: contig of 821 bp in length
* * gap of unknown length
* * 29881: contig of 807 bp in length
* * gap of unknown length
* * 30688: contig of 792 bp in length
* * gap of unknown length
* * 31480: contig of 783 bp in length
* * gap of unknown length
* * 32263: contig of 822 bp in length
* * gap of unknown length
* * 33085: contig of 816 bp in length
* * gap of unknown length
* * 33901: contig of 791 bp in length
* * gap of unknown length
* * 34692: contig of 773 bp in length
* * gap of unknown length
* * 35465: contig of 814 bp in length
* * gap of unknown length
* * 36279: contig of 807 bp in length
* * gap of unknown length
* * 37086: contig of 784 bp in length
* * gap of unknown length
* * 37870: contig of 777 bp in length
* * gap of unknown length
* * 38647: contig of 757 bp in length
* * gap of unknown length
* * 39404: contig of 785 bp in length
* * gap of unknown length
* * 40189: contig of 816 bp in length
* * gap of unknown length
* * 41005: contig of 821 bp in length
* * gap of unknown length
* * 41826: contig of 825 bp in length
* * gap of unknown length
* * 42651: contig of 820 bp in length
* * gap of unknown length
* * 43471: contig of 720 bp in length
* * gap of unknown length
* * 44191: contig of 800 bp in length
* * gap of unknown length
* * 44991: contig of 809 bp in length
* * gap of unknown length
* * 45800: contig of 805 bp in length
* * gap of unknown length
* * 46605: contig of 786 bp in length
* * gap of unknown length
```

```
* * 47391 48179: contig of 789 bp in length
* * gap of unknown length
* * 48180 48961: contig of 782 bp in length
* * gap of unknown length
* * 48962 49828: contig of 867 bp in length
* * gap of unknown length
* * 49829 50651: contig of 823 bp in length
* * gap of unknown length
* * 50652 51468: contig of 817 bp in length
* * gap of unknown length
* * 51469 52279: contig of 811 bp in length
* * gap of unknown length
* * 52280 53098: contig of 819 bp in length
* * gap of unknown length
* * 53099 53910: contig of 812 bp in length
* * gap of unknown length
* * 53911 54696: contig of 786 bp in length
* * gap of unknown length
* * 54697 55480: contig of 784 bp in length
* * gap of unknown length
* * 55481 56288: contig of 808 bp in length
* * gap of unknown length
* * 56289 57119: contig of 831 bp in length
* * gap of unknown length
* * 57120 57941: contig of 822 bp in length
* * gap of unknown length
* * 57942 58742: contig of 801 bp in length
* * gap of unknown length
* * 58743 59556: contig of 814 bp in length
* * gap of unknown length
* * 59557 60366: contig of 810 bp in length.
FEATURES
      Location/Qualifiers
      1..60366
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="2_K12"
      /clone_lib="RPC1-11 Human Male BAC"
```

```
Query Match      2.5%; Score 21; DB 43; Length 60366;
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 787 cttttcccaataaaaaaaa 807
      |||||
DB 11989 CTTTCCCAATAAAAAAAA 11969
```

```
RESULT 9
CEY113B8_1
WPCOMMENT
```

```
Sequence split into 4 fragments LOCUS CEY113B8 Accession AL021573
Fragment Name      Begin      End
CEY113B8_0          1      110000
CEY113B8_1         100001     210000
CEY113B8_2         200001     310000
CEY113B8_3         300001     357994
Continuation (2 of 4) of CEY113B8 from base 100001 (AL021573 Caenorhabditis elegans c
```

```
Query Match      2.5%; Score 21; DB 32; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.88;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 797 taaaaaaaatcatcccaaa 817
      |||||
DB 88629 TAAAAAATCATCCAAA 88649
```

```
RESULT 10
CEY113B8_2
WPCOMMENT
```

```
Sequence split into 4 fragments LOCUS CEY113B8 Accession AL021573
Fragment Name      Begin      End
```

```

CEY113B8_0      1      110000
CEY113B8_1      100001  210000
CEY113B8_2      200001  310000
CEY113B8_3      300001  357994
Continuation (3 of 4) of CEY113B8 from base 200001 (AL021573 Caenorhabditis elegans chr

Query Match      2.5%; Score 21; DB 32; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.88;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 797 taataaaaaaatcatcaaaa 817
|||||
Db 86545 TAAAAAAAATCATCAAAA 86565

RESULT 11
AC010178      163678 bp      DNA      HTG      16-SEP-1999
LOCUS      Homo sapiens clone RPC11-30309, *** SEQUENCING IN PROGRESS ***, 42
DEFINITION      unordered pieces.
ACCESSION      AC010178
VERSION      AC010178.1 GI:5882328
KEYWORDS      HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS      Muzny, D.M., Adams, C., Aronson, A., Bailey, M., Barberia, J.,
Blankenburg, K., Bodota, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C.,
Bunac, C., Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z.,
Cox, C., David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,
Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,
Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,
Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hoques, M.,
Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,
Kelly, S., Kondrjewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,
Lichtarge, O., Liu, J., Liu, W., Logan, O., Lu, J., Lucier, R.,
Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M., Morris, S.,
Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Oswal, G.,
Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M.,
Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, E.,
Shen, H., Simon, M., Sparks, A., Stamps, A., Suckang, R., Taber, P.,
Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wambah, M., Watlington, S.,
Weinstock, G., Weinstock, I.R., Williamson, A., Worley, K., Wren, J.,
Wrensford, G., Yu, W., Zhou, X. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 163678)
Worley, K.C.
Direct Submission
TITLE      Submitted (15-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 42 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 13971: contig of 13971 bp in length
* 13972 13991: gap of unknown length
* 13992 27643: contig of 13652 bp in length
* 27644 27663: gap of unknown length
* 27664 39058: contig of 11395 bp in length
* 39059 39078: gap of unknown length
* 39079 48651: contig of 9573 bp in length
* 48652 48671: gap of unknown length
* 48672 55565: contig of 6894 bp in length
* 55566 55585: gap of unknown length
61503: contig of 5918 bp in length
61523: gap of unknown length
61524 67253: contig of 5730 bp in length
67254 67273: gap of unknown length
67274 72990: contig of 5717 bp in length
72991 73010: gap of unknown length
73011 78484: contig of 5474 bp in length
78485 78504: gap of unknown length
78505 83670: contig of 5166 bp in length
83671 83690: gap of unknown length
83691 87911: contig of 4221 bp in length
87912 87931: gap of unknown length
87932 92091: contig of 4160 bp in length
92092 92111: gap of unknown length
92112 96138: contig of 4027 bp in length
96139 96158: gap of unknown length
100058: contig of 3900 bp in length
100059 100078: gap of unknown length
100079 103934: contig of 3856 bp in length
103935 103954: gap of unknown length
103955 107641: contig of 3687 bp in length
107642 107661: gap of unknown length
107662 111289: contig of 3628 bp in length
111290 111309: gap of unknown length
111310 114916: contig of 3607 bp in length
114917 114936: gap of unknown length
114937 118272: contig of 3336 bp in length
118273 118292: gap of unknown length
118293 121583: contig of 3291 bp in length
121584 121603: gap of unknown length
121604 124787: contig of 3184 bp in length
124788 124807: gap of unknown length
124808 127934: contig of 3127 bp in length
127935 127954: gap of unknown length
127955 130950: contig of 2996 bp in length
130951 130970: gap of unknown length
130971 133664: contig of 2694 bp in length
133665 133684: gap of unknown length
133685 136112: contig of 2428 bp in length
136113 136132: gap of unknown length
136133 138522: contig of 2390 bp in length
138523 138541: gap of unknown length
140888 140888: contig of 2346 bp in length
140889 140908: gap of unknown length
140909 143091: contig of 2183 bp in length
143092 143111: gap of unknown length
143112 145206: contig of 2095 bp in length
145207 145226: gap of unknown length
145227 147128: contig of 1902 bp in length
147129 147148: gap of unknown length
147149 148895: contig of 1747 bp in length
148896 148915: gap of unknown length
148916 150652: contig of 1737 bp in length
150653 150672: gap of unknown length
150673 152407: contig of 1735 bp in length
152408 152427: gap of unknown length
152428 154126: contig of 1699 bp in length
154127 154146: gap of unknown length
154147 155693: contig of 1547 bp in length
155694 155713: gap of unknown length
155714 157164: contig of 1451 bp in length
157165 157184: gap of unknown length
157185 158582: contig of 1398 bp in length
158583 158602: gap of unknown length
158603 159953: contig of 1351 bp in length
159954 159973: gap of unknown length
161202 161202: contig of 1229 bp in length
161203 161222: gap of unknown length
161223 162397: contig of 1175 bp in length
162398 162417: gap of unknown length
162418 163472: contig of 1055 bp in length
163473 163492: gap of unknown length
163493 163678: contig of 186 bp in length.

```

```

FEATURES
  source
    Location/Qualifiers
      1..163678
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="RPCII-30309"
BASE COUNT 39743 a 40292 c 40916 g 41669 t 1058 others
ORIGIN

Query Match
Best Local Similarity 2.5%; Score 21; DB 41; Length 163678;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 786 tctttccaataaaaaaaa 806
|||||
DB 98953 TCTTTCCAAATAAAAAAAA 98973

RESULT 12
AC006735
LOCUS AC006735 166214 bp DNA HTG 25-FEB-1999
DEFINITION Caenorhabditis elegans clone Y34D9, *** SEQUENCING IN PROGRESS ***
ACCESSION AC006735
VERSION AC006735.3 GI:4309801
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE
  AUTHORS Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
  TITLE Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
  JOURNAL 1 (bases 1 to 166214)
  AUTHORS Waterston,R.H.
  TITLE The sequence of Caenorhabditis elegans clone
  JOURNAL 2 (bases 1 to 166214)
  AUTHORS Waterston,R.H.
  TITLE Unpublished
  JOURNAL 2 (bases 1 to 166214)
  AUTHORS Waterston,R.H.
  TITLE Direct Submission
  JOURNAL Submitted (23-FEB-1999) Genome Sequencing Center, Washington
  MO 63108, USA
COMMENT
  On Mar 1, 1999 this sequence version replaced gi:4263429.
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 4 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  * 1 2055: contig of 2055 bp in length
  * 2056 2069: gap of unknown length
  * 2070 20299: contig of 18230 bp in length
  * 20300 20313: gap of unknown length
  * 20314 70328: contig of 50015 bp in length
  * 70329 70342: gap of unknown length
  * 70343 166214: contig of 95872 bp in length.
FEATURES
  source
    Location/Qualifiers
      1..166214
        /organism="Caenorhabditis elegans"
        /db_xref="taxon:6239"
        /clone="Y34D9"
BASE COUNT 52633 a 31045 c 31031 g 51463 t 42 others
ORIGIN

Query Match
Best Local Similarity 2.5%; Score 21; DB 33; Length 166214;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 788 ttttccaataaaaaaaaat 808
|||||
DB 93214 TTTTCCAAATAAAAAAAAAT 93234

FEATURES
  source
    Location/Qualifiers
      1..163678
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="RPCII-30309"
BASE COUNT 39743 a 40292 c 40916 g 41669 t 1058 others
ORIGIN

Query Match
Best Local Similarity 2.5%; Score 21; DB 44; Length 176586;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 790 ttccaataataaaaaaatca 810
|||||
DB 35711 TTCCAATAATAAAAAAATCA 35691

RESULT 14
AC007799/c
LOCUS AC007799 185000 bp DNA HTG 16-NOV-1999
DEFINITION Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***
ACCESSION AC007799
VERSION AC007799.4 GI:6437515
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
  AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  TITLE Eutheria; Primates; Catarrhini; Hominidae; Homo.
  JOURNAL 1 (bases 1 to 185000)
  AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
  TITLE Direct Submission
  JOURNAL Unpublished
  REFERENCE 2 (bases 1 to 185000)
  AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
  TITLE Direct Submission

```

JOURNAL Submitted (12-JUN-1999) Department of Genetics, Stanford Human Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA  
 COMMENT On Nov 16, 1999 this sequence version replaced gi:6017090.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 34 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1316: contig of 1316 bp in length  
 \* 1317 1685: gap of unknown length  
 \* 1686 3021: contig of 1336 bp in length  
 \* 3022 3390: gap of unknown length  
 \* 3391 5853: contig of 2463 bp in length  
 \* 5854 6222: gap of unknown length  
 \* 6223 7800: contig of 1578 bp in length  
 \* 7801 8169: gap of unknown length  
 \* 8170 9931: contig of 1762 bp in length  
 \* 9932 10300: gap of unknown length  
 \* 10301 11649: contig of 1349 bp in length  
 \* 11650 12018: gap of unknown length  
 \* 12019 14568: contig of 2550 bp in length  
 \* 14569 14937: gap of unknown length  
 \* 14938 16223: contig of 1286 bp in length  
 \* 16224 16592: gap of unknown length  
 \* 16593 18641: contig of 2049 bp in length  
 \* 18642 19009: gap of unknown length  
 \* 19010 20899: contig of 1890 bp in length  
 \* 20900 21267: gap of unknown length  
 \* 21268 22488: contig of 1221 bp in length  
 \* 22489 22856: gap of unknown length  
 \* 22857 24907: contig of 2051 bp in length  
 \* 24908 25275: gap of unknown length  
 \* 25276 27361: contig of 2086 bp in length  
 \* 27362 27729: gap of unknown length  
 \* 27730 29489: contig of 1760 bp in length  
 \* 29490 29857: gap of unknown length  
 \* 29858 33323: contig of 3466 bp in length  
 \* 33324 33691: gap of unknown length  
 \* 33692 35072: contig of 1381 bp in length  
 \* 35073 35440: gap of unknown length  
 \* 35441 38098: contig of 2638 bp in length  
 \* 38099 38466: gap of unknown length  
 \* 38467 42473: contig of 4007 bp in length  
 \* 42474 42841: gap of unknown length  
 \* 42842 45566: contig of 2725 bp in length  
 \* 45567 45934: gap of unknown length  
 \* 45935 49948: contig of 4014 bp in length  
 \* 49949 50316: gap of unknown length  
 \* 50317 51878: contig of 1562 bp in length  
 \* 51879 52246: gap of unknown length  
 \* 52247 57059: contig of 4813 bp in length  
 \* 57060 57437: gap of unknown length  
 \* 57438 59702: contig of 2275 bp in length  
 \* 59703 60070: gap of unknown length  
 \* 60071 64068: contig of 3998 bp in length  
 \* 64069 64436: gap of unknown length  
 \* 64437 68196: contig of 3760 bp in length  
 \* 68197 68564: gap of unknown length  
 \* 68565 72554: contig of 3990 bp in length  
 \* 72555 72922: gap of unknown length  
 \* 72923 80178: contig of 7256 bp in length  
 \* 80179 80546: gap of unknown length  
 \* 80547 88065: contig of 7519 bp in length  
 \* 88066 88433: gap of unknown length  
 \* 88434 100937: contig of 12564 bp in length  
 \* 100938 101365: gap of unknown length  
 \* 101366 109749: contig of 8384 bp in length  
 \* 109750 110117: gap of unknown length  
 \* 110118 123838: contig of 13721 bp in length  
 \* 123839 124206: gap of unknown length

\* 124207 141747: contig of 17541 bp in length  
 \* 141748 142115: gap of unknown length  
 \* 142116 166417: contig of 24302 bp in length  
 \* 166418 166785: gap of unknown length  
 \* 166786 185000: contig of 18215 bp in length.  
 FEATURES Location/Qualifiers  
 source 1..185000  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 BASE COUNT 50887 a 33471 c 33515 g 54945 t 12182 others  
 ORIGIN

Query Match 2.5%; Score 21; DB 43; Length 185000;  
 Best Local Similarity 100.0%; Pred. No. 0.91;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 786 tctttccaaataaaaaaa 806  
 ||||||||||||||||  
 Db 33316 TCTTTTCCAAATAAAAAA 33296

RESULT 15  
 CEY7A9 224129 bp DNA HTG 03-DEC-1998  
 LOCUS Caenorhabditis elegans chromosome IV clone Y7A9, \*\*\* SEQUENCING IN  
 DEFINITION PROGRESS \*\*\*, in unordered pieces.  
 SOURCE 295312  
 VERSION 295312.1 GI:3355732  
 KEYWORDS HTG; HTGS\_PHASE1.  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
 Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 224129)  
 AUTHORS McMurray, A.  
 DIRECT SUBMISSION  
 TITLE Submitted (03-DEC-1998) Nematode Sequencing Project, Sanger Centre,  
 Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,  
 Washington University, St. Louis, MO 63110, USA. E-mail:  
 jes@sanger.ac.uk or rvnematode.wustl.edu  
 COMMENT On Jul 30, 1998 this sequence version replaced gi:2528903.  
 Order of segments is not known; 800 n's separate segments.  
 IMPORTANT: This sequence is unfinished and does not necessarily  
 represent the correct sequence. Work on the sequence is in progress  
 and the release of this data is based on the understanding that the  
 sequence may change as work continues. The sequence may be  
 contaminated with foreign sequence from E.coli, yeast, vector,  
 phage etc.

\* NOTE: This is a 'working draft' sequence.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 FEATURES Location/Qualifiers  
 source 1..224129  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /chromosome="IV"  
 /clone="Y7A9"  
 BASE COUNT 72928 a 37724 c 37595 g 73444 t 2438 others  
 ORIGIN

Query Match 2.5%; Score 21; DB 32; Length 224129;  
 Best Local Similarity 100.0%; Pred. No. 0.92;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 797 taaaaaaaatcatccaaaa 817  
 ||||||||||||||||  
 Db 155900 TAAAAAAAATCATCCAAA 155920



Search completed: May 1, 2000, 17:01:56  
Job time: 21816 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 18:54:09 ; Search time 250.34 Seconds  
(without alignments)  
825.512 Million cell updates/sec

Title: US-09-215-435-124

Perfect score: 826

Sequence: 1 aacagagggtgcccattgggtt.....atcatccaaaaa 826

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 311585 seqs, 125096042 residues

Word size : 0

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	275	33.3	447	1 X40582	Human secreted pro
2	19	2.3	1227	1 V42549	Mouse decitin-2 cDN
3	19	2.3	2050	1 N80749	Sequence of bovine
4	19	2.3	2050	1 Q23215	Bovine interleukin
5	19	2.3	2104	1 V40506	Human secreted pro
6	19	2.3	3102	1 V17260	Human protein p164
7	19	2.3	4855	1 V22252	Human myosin light
8	19	2.3	110000	1 X20248_00	Borrelia burgdorfe
9	18	2.2	456	1 X00632	Human secreted pro
10	18	2.2	509	1 V34180	Human secreted pro
11	18	2.2	572	1 V44301	Human secreted pro
12	18	2.2	653	1 V59522	Human secreted pro
13	18	2.2	1113	1 Q46775	Maize dwarf mosaic
14	18	2.2	1348	1 Q78278	T. niveum GAPDH. N
15	18	2.2	1355	1 N80457	Sequence of cDNA e
16	18	2.2	1357	1 X51739	DNA encoding a hum
17	18	2.2	1485	1 X51523	cDNA encoding huma
18	18	2.2	2045	1 V34297	Human secreted pro
19	18	2.2	2328	1 N81399	Human secreted pro
20	18	2.2	2331	1 N81116	L-phenylalanine am
21	18	2.2	2331	1 N81101	Sequence encoding
22	18	2.2	2640	1 Q26664	bDAT. cDNA encodin
23	18	2.2	2783	1 Q34758	Nucleotide sequenc
24	18	2.2	2783	1 Q4067	N. crassa mtr gene
25	18	2.2	7750	1 X13153	Enterococcus faeca
26	17	2.1	21	1 Q75613	Reverse transcript
27	17	2.1	112	1 T25481	Human gene signatu
28	17	2.1	179	1 V16146	Microsatellite mar
29	17	2.1	207	1 Q50399	GPF - 1E, part of
30	17	2.1	228	1 Q60633	Human brain expres
31	17	2.1	251	1 X10920	Human biallelic po
32	17	2.1	260	1 T34469	NP4 fusion cDNA. C
33	17	2.1	260	1 T28175	Monoclonal antibod

34 17 2.1 274 1 V32425 Homo sapiens clone  
35 17 2.1 379 1 Q86302 Wilson disease gen  
36 17 2.1 461 1 V89657 EST clone DA10. Ne  
37 17 2.1 517 1 Q13621 Derf II gene in pf  
38 17 2.1 560 1 T14356 Nsk2 receptor alte  
39 17 2.1 560 1 V38550 Nsk2 isoform cDNA  
40 17 2.1 581 1 Q14657 Bay thioesterase g  
41 17 2.1 599 1 V84500 Human secreted pro  
42 17 2.1 687 1 X30410 DNA encoding a hum  
43 17 2.1 735 1 V29153 Mouse zins3 gene.  
44 17 2.1 771 1 V10247 Human herpesvirus  
45 17 2.1 840 1 V84447 Human secreted pro

#### ALIGNMENTS

##### RESULT 1

X40582  
ID X40582 standard; cDNA; 447 BP.  
AC X40582;  
DT 18-JUN-1999 (first entry)  
DE Human secreted protein 5', EST SEQ ID No: 182.  
KW Human; secreted protein; EST: expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
OS Homo sapiens.  
PN WO9906550-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; IB1232.  
PR 01-AUG-1997; US-905144.  
PA (GEST) GENSET.  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
DR WPI: 99-153780/13.  
P-PSDB: Y11860.  
PT New isolated prostate-derived nucleic acids - used to develop  
PT products which may have cytokine, immune regulatory, haematopoiesis  
PT regulating, anti-inflammatory or tumour inhibition activity  
PS Claim 1: Page 285: 675pp: English.  
CC X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins expressed in prostate, and encode the proteins given in  
CC Y11716 to Y11993 respectively. The proteins given represent the signal  
CC peptide and an N-terminal fragment of a secreted protein. The nucleic  
CC acid sequences can be used for producing secreted human gene products.  
CC They can also be used to develop products for diagnosis and therapy. The  
CC proteins obtained may have cytokine activity, cell proliferation and  
CC differentiation activity, haematopoiesis regulating activity, tissue  
CC growth regulating activity, reproductive hormone regulating activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,  
CC receptor/ligand activity, anti-inflammatory activity, tumour inhibition  
CC activity or other activities. The products can be used in forensic, gene  
CC therapy and chromosome mapping procedures. The sequences can also be used  
CC for obtaining corresponding promoter sequences. The nucleic acids  
CC encoding the signal peptides can be used for directing extracellular  
CC secretion of a polypeptide or the insertion of a polypeptide into a  
CC membrane, or importing a polypeptide into a cell.  
SQ Sequence 447 BP; 98 A; 110 C; 136 G; 99 T;

Query Match 33.3%; Score 275; DB 1; Length 447;  
Best Local Similarity 99.5%; Pred. No. 4.2e-114;  
Matches 375; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacagagggtgcccattgggttggacaatgagctgggtcacagcagcactgttactgggtc 60

Db 71 AACACAGAGGTGCCCATGGGTTGGACATGAGGCTGGTCCACAGCAGCAGTGTACTGGTGC 130

Qy 61 catatggtggtcactgagcagaggtgagacacccctgtgcccattagccctcc 120

Db 131 TCATGATGTTGTCTACTGGAGCAGGAGTGAAGACAGCCCGTGTGCCCATGAGGCCCTCT 190



```
PR 23-DEC-1991; US-813235.
PA (IMMV ) IMMUNEX CORP.
PI Cerretti DP, Maliszewski CR, Schoenborn M;
DR WPI; 93-395281/49.
PT P-PSDB; R43900.
PT Homogeneous bovine interleukin-1 alpha - used as a vaccine
PT adjuvant to potentiate the immune response to an antigen or to
PT promote wound healing
PS Disclosure; Fig 1; 10pp; English.
CC The sequence is that encoding bovine interleukin-1 alpha (bIL-1alpha).
SQ Sequence 2050 BP; 623 A; 440 C; 409 G; 578 T;

Query Match 2.3%; Score 19; DB 1; Length 2050;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 cagcctaccaggtccctc 433
Db 1526 CAGCCTACCAGGCTCCCTC 1544
|||||

RESULT 5
ID V40506 standard; cDNA; 2104 BP.
AC V40506;
DE 09-NOV-1998 (first entry)
DE Human secreted protein DF518_3 cDNA.
KW Secreted protein; DF518_3; human; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 3..1940
FT /tag= a
FT WO9830582-A2.
PD 16-JUL-1998.
PN 08-JAN-1998; U00289.
PR 08-JAN-1998; US-004680.
PR 09-JAN-1997; US-780890.
PA (GEMV ) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 98-413681/35.
DR P-PSDB; W29653.
PT New isolated nucleic acids and secreted proteins - obtained from
PT human foetal kidney, human adult retina, human foetal brain, human
PT adult brain and human adult blood cDNA libraries
PS Claim 28; Page 73-74; 103pp; English.
CC This cDNA clone, designated DF518_3, includes a coding region for a
CC 645-amino acid secreted protein (see W29653).. Full-length DF518_3
CC cDNA was isolated from a human adult brain cDNA library using
CC methods which are selective for cDNAs encoding secreted proteins,
CC or was identified as encoding a secreted or transmembrane protein
CC on the basis of the amino acid sequence of the encoded protein.
CC The invention relates to 9 cDNA clones (see V40501-09), all
CC deposited as ATCC 98290, that code for human secreted proteins
CC (see W29648-56) of the foetal kidney or brain, or adult retina,
CC brain or blood. Host cells and methods of producing mature
CC polypeptides are also claimed. The polynucleotides and polypeptides
CC can be used as e.g. nutritional sources or supplements or may
CC exhibit e.g. cytokine and cell proliferation or differentiation
CC activity, immunostimulant or immunosuppressive activity,
CC haematopoiesis regulating activity, receptor/ligand activity,
CC antiinflammatory activity, cadherin or tumour invasion suppressor
CC activity, tumour inhibition activity or other activities. DF518_3
CC demonstrates some similarity to a number of database sequences.
SQ Sequence 2104 BP; 566 A; 582 C; 616 G; 340 T;

Query Match 2.3%; Score 19; DB 1; Length 2104;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 808 tcattcaaaataaaaaaa 826
|||||

PR 23-DEC-1991; US-813235.
PA (IMMV ) IMMUNEX CORP.
PI Cerretti DP, Maliszewski CR, Schoenborn M;
DR WPI; 93-395281/49.
PT P-PSDB; R43900.
PT Homogeneous bovine interleukin-1 alpha - used as a vaccine
PT adjuvant to potentiate the immune response to an antigen or to
PT promote wound healing
PS Disclosure; Fig 1; 10pp; English.
CC The sequence is that encoding bovine interleukin-1 alpha (bIL-1alpha).
SQ Sequence 2050 BP; 623 A; 440 C; 409 G; 578 T;

Query Match 2.3%; Score 19; DB 1; Length 3102;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 789 ttcccaataaaaaaaa 807
Db 3083 TTTCCAATAAAAAAAA 3101
|||||

RESULT 6
ID V17260 standard; cDNA; 3102 BP.
AC V17260;
DE 01-JUN-1998 (first entry)
DE Human protein p164 coding sequence.
KW Human; protein p164; Rho protein; myosin combined subunit;
KW combining activity; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..3093
FT /tag= a
FT J10057074-A.
PD 03-MAR-1998.
PF 11-JUL-1996; 201325.
PR 05-JAN-1996; JP-017151.
PR 20-NOV-1995; JP-325130.
PR 05-DEC-1995; JP-344605.
PA (KIRI ) KIRIN BREWERY KK.
DR WPI; 98-210406/19.
DR P-PSDB; W41378.
PT New physiologically active protein p138 - useful for, e.g. screening
PT for gene therapy for treatment of disease related to Rho protein
PS Example 10; Page 28-32; 36pp; Japanese.
CC This sequence represents the coding sequence for the human protein p164,
CC which is a protein of the invention. The protein has the combining
CC activity of an active type Rho protein. The protein can also contain a
CC phosphorylated site. The DNA sequence encoding the protein can be used
CC to screen for a genetic drug useful for treating a disease related to
CC Rho protein. The protein or a peptide can be used for screening for an
CC inhibitor that inhibits the combination of active Rho protein and myosin
CC combined subunit.
SQ Sequence 3102 BP; 1153 A; 589 C; 710 G; 650 T;

Query Match 2.3%; Score 19; DB 1; Length 3102;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 789 ttcccaataaaaaaaa 807
Db 3083 TTTCCAATAAAAAAAA 3101
|||||

RESULT 7
ID V22252 standard; cDNA; 4855 BP.
AC V22252;
DE 07-JUL-1998 (first entry)
DE Human myosin light chain bound subunit cDNA.
KW Human; myosin light chain; bound subunit; Rho type protein;
KW phosphatase promoter; tumour diagnosis; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 243..3335
FT /tag= a
FT J10077299-A.
PD 24-MAR-1998.
PF 28-FEB-1997; 061847.
PR 11-JUL-1996; JP-201325.
PA (KIRI ) KIRIN BREWERY KK.
DR WPI; 98-264599/24.
DR P-PSDB; W53572.
PT Protein from human myosin light chain gene - useful for tumour
PT diagnosis
PS Claim 13; Pages 30-35; 47pp; Japanese.
CC The present sequence encodes a human myosin light chain bound
CC subunit, which binds Rho type protein, promotes myosin light chain
CC phosphatase catalytic subunit activity, has its gene positioned at
```

CC human chromosome 12q15-21.2 and has a molecular weight of 130 to  
CC 135 kDa by SDS-PAGE. The gene encoding the protein can be used for  
CC the diagnosis of tumours.  
SQ Sequence 4855 BP; 1616 A; 922 C; 1036 G; 1281 T;

Query Match 2.3%; Score 19; DB 1; Length 4855;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 789 ttctccaaataaataaaaaa 807  
|||||  
DB 3325 TTTCCAATAAATAAAAAA 3343

RESULT 8  
X20248\_00  
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
WP Fragment Name Begin End  
WP X20248\_00 1 110000  
WP X20248\_01 100001 210000  
WP X20248\_02 200001 310000  
WP X20248\_03 300001 410000  
WP X20248\_04 400001 510000  
WP X20248\_05 500001 610000  
WP X20248\_06 600001 710000  
WP X20248\_07 700001 810000  
WP X20248\_08 800001 910000  
WP X20248\_09 900001 910715  
ID X20248 standard; DNA; 910715 BP.  
AC X20248  
DT 04-MAY-1999 (first entry)  
DE Borrelia burgdorferi polynucleotide sequence #1.  
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;  
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;  
KW infection; diagnosis; characterisation; detection; ds.  
OS Borrelia burgdorferi.  
PN W09858943-A1.  
PD 30-DEC-1998.  
PF 18-JUN-1998; U12764.  
PR 03-SEP-1997; US-057483.  
PR 20-JUN-1997; US-050359.  
PR 22-JUL-1997; US-053344.  
PR 22-JUL-1997; US-053377.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA (MEDI-) MEDIMUNE INC.  
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,  
PI White OR;  
PI WPI; 99-081217/07.  
PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
PT products for the detection, diagnosis, characterisation, prevention  
PT and therapy of infections, particularly Lyme disease  
PS Claim 1; Page 157-671; 1128pp; English.  
CC X20248 to X20402 represent polynucleotide sequences isolated from  
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for  
CC the detection, diagnosis, characterisation, prevention and therapy of  
CC Bb infections, e.g. Lyme disease. They can also be used for the  
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs  
CC to a family of motile, spiral-shaped bacteria called Spirochetes.  
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and  
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as  
CC Lyme disease.  
SQ Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T;

Query Match 2.3%; Score 19; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 786 tctttccaaataaataaaaaa 804  
|||||  
DB 45636 TCCTTTCCAAATAAATAAAAAA 45654

RESULT 9  
X00632  
ID X00632 standard; DNA; 456 BP.  
AC X00632;  
DT 25-MAR-1999 (first entry)  
DE Human secreted protein gene 22 clone HFEAF41.  
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
KW immune system; asthma; lymphocytic disease; brain; hepatitis; lymphoma;  
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
OS Homo sapiens.  
PN W09842738-A1.  
PD 01-OCT-1998.  
PF 19-MAR-1998; U05311.  
PR 30-MAY-1997; US-050937.  
PR 21-MAR-1997; US-041276.  
PR 21-MAR-1997; US-041277.  
PR 21-MAR-1997; US-041281.  
PR 21-MAR-1997; US-042344.  
PR 30-MAY-1997; US-048069.  
PR 30-MAY-1997; US-048094.  
PR 30-MAY-1997; US-048095.  
PR 30-MAY-1997; US-048096.  
PR 30-MAY-1997; US-048099.  
PR 30-MAY-1997; US-048131.  
PR 30-MAY-1997; US-048135.  
PR 30-MAY-1997; US-048154.  
PR 30-MAY-1997; US-048160.  
PR 30-MAY-1997; US-048186.  
PR 30-MAY-1997; US-048187.  
PR 30-MAY-1997; US-048188.  
PR 30-MAY-1997; US-048350.  
PR 30-MAY-1997; US-048351.  
PR 30-MAY-1997; US-048352.  
PR 30-MAY-1997; US-048355.  
PR 05-AUG-1997; US-054804.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA Brewer LA, Duan R, Ehner R, Ferrie AM, Florence KA,  
PI Greene JM, Hu JS, Larleir DW, Moore PA, NI J, Olsen HS,  
PI Rosen CA, Ruben SM, Shi Y, Young P;  
DR WPI; 99-070066/06.  
DR P-PSDB; W67828.  
PT New isolated human genes and the secreted polypeptides they encode -  
PT useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders  
PS Claim 1; Page 188; 385pp; English.  
CC This sequence represents a nucleic acid molecule which encodes a secreted  
CC human protein. The gene number, and the clone it is derived from, are  
CC detailed in the descriptor line. The gene can be used to generate fusion  
CC proteins by linking to the gene to a human immunoglobulin Fc portion  
CC (e.g. X00602) for increasing the stability of the fused protein as  
CC compared to the human protein only.  
CC The invention relates to 87 novel genes and their fragments (nucleic acid  
CC sequences: X00611-X00724; amino acid sequences W67807-W68004) which  
CC are useful for preventing, treating or ameliorating medical conditions  
CC e.g. by protein or gene therapy. Also, pathological conditions can be  
CC diagnosed by determining the amount of the new polypeptides in a sample  
CC or by determining the presence of mutations in the new polynucleotides.  
CC Specific uses are described for each of the 87 polynucleotides, based on  
CC which tissues they are most highly expressed in (see X00611 for described  
CC uses).  
SQ Sequence 456 BP; 117 A; 136 C; 104 G; 98 T;

Query Match 2.2%; Score 18; DB 1; Length 456;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 809 catcaaaaaaaaaaaaaa 826  
 Db 416 CATCCAAAAA 433

RESULT 10

V34180  
 ID V34180 standard; DNA; 509 BP.  
 AC V34180;  
 DT 28-JAN-1999 (first entry)  
 DE Human secreted protein gene 27 clone HSPAF93.  
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumor; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; fetal deficiency; blood; allergy; renal; ds;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 OS Homo sapiens.  
 PN W09839446-A2.  
 PD 11-SEP-1998.  
 PF 06-MAR-1998. U04492.  
 PR 07-MAR-1997; US-038621.  
 PR 07-MAR-1997; US-040161.  
 PR 07-MAR-1997; US-040162.  
 PR 07-MAR-1997; US-040163.  
 PR 07-MAR-1997; US-040333.  
 PR 07-MAR-1997; US-040334.  
 PR 07-MAR-1997; US-040336.  
 PR 07-MAR-1997; US-040826.  
 PR 11-APR-1997; US-043311.  
 PR 11-APR-1997; US-043312.  
 PR 11-APR-1997; US-043313.  
 PR 11-APR-1997; US-043314.  
 PR 11-APR-1997; US-043315.  
 PR 11-APR-1997; US-043568.  
 PR 11-APR-1997; US-043569.  
 PR 11-APR-1997; US-043576.  
 PR 11-APR-1997; US-043578.  
 PR 11-APR-1997; US-043580.  
 PR 11-APR-1997; US-043659.  
 PR 11-APR-1997; US-043670.  
 PR 11-APR-1997; US-043671.  
 PR 11-APR-1997; US-043672.  
 PR 11-APR-1997; US-043674.  
 PR 23-MAY-1997; US-047492.  
 PR 23-MAY-1997; US-047500.  
 PR 23-MAY-1997; US-047501.  
 PR 23-MAY-1997; US-047502.  
 PR 23-MAY-1997; US-047503.  
 PR 23-MAY-1997; US-047581.  
 PR 23-MAY-1997; US-047582.  
 PR 23-MAY-1997; US-047583.  
 PR 23-MAY-1997; US-047584.  
 PR 23-MAY-1997; US-047585.  
 PR 23-MAY-1997; US-047586.  
 PR 23-MAY-1997; US-047587.  
 PR 23-MAY-1997; US-047588.  
 PR 23-MAY-1997; US-047589.  
 PR 23-MAY-1997; US-047590.  
 PR 23-MAY-1997; US-047592.  
 PR 23-MAY-1997; US-047593.  
 PR 23-MAY-1997; US-047594.  
 PR 23-MAY-1997; US-047595.  
 PR 23-MAY-1997; US-047596.  
 PR 23-MAY-1997; US-047597.  
 PR 23-MAY-1997; US-047598.  
 PR 23-MAY-1997; US-047599.  
 PR 23-MAY-1997; US-047600.  
 PR 23-MAY-1997; US-047601.  
 PR 23-MAY-1997; US-047612.  
 PR 23-MAY-1997; US-047613.

PR 23-MAY-1997; US-047614.  
 PR 23-MAY-1997; US-047615.  
 PR 23-MAY-1997; US-047617.  
 PR 23-MAY-1997; US-047618.  
 PR 23-MAY-1997; US-047632.  
 PR 23-MAY-1997; US-047633.  
 PR 06-JUN-1997; US-048964.  
 PR 06-JUN-1997; US-048974.  
 PR 22-AUG-1997; US-056630.  
 PR 22-AUG-1997; US-056631.  
 PR 22-AUG-1997; US-056632.  
 PR 22-AUG-1997; US-056636.  
 PR 22-AUG-1997; US-056637.  
 PR 22-AUG-1997; US-056662.  
 PR 22-AUG-1997; US-056664.  
 PR 22-AUG-1997; US-056845.  
 PR 22-AUG-1997; US-056862.  
 PR 22-AUG-1997; US-056864.  
 PR 22-AUG-1997; US-056866.  
 PR 22-AUG-1997; US-056887.  
 PR 22-AUG-1997; US-056888.  
 PR 22-AUG-1997; US-056889.  
 PR 22-AUG-1997; US-056892.  
 PR 22-AUG-1997; US-056893.  
 PR 22-AUG-1997; US-056894.  
 PR 22-AUG-1997; US-056903.  
 PR 22-AUG-1997; US-056908.  
 PR 22-AUG-1997; US-056909.  
 PR 22-AUG-1997; US-056910.  
 PR 22-AUG-1997; US-056911.  
 PR 05-SEP-1997; US-057650.  
 PR 05-SEP-1997; US-057761.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,  
 PI Feng P, Ferrie AM, Fischer CL, Graves KA, Greene JM, Hu JS,  
 PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,  
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;  
 DR WPI; 98-609887/51.  
 DR P-ESDB; W75083.  
 PT New isolated human genes and the secreted polypeptides they encode  
 PT - useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 PS Claim 1; Page 183; 447pp; English.  
 CC This sequence represents a nucleic acid molecule which encodes a secreted  
 CC human protein. The gene number, and the clone it is derived from, are  
 CC detailed in the descriptor line. The gene can be used to generate fusion  
 CC proteins by linking to the gene to a human immunoglobulin Fc portion  
 CC (e.g. V34145) for increasing the stability of the fused protein as  
 CC compared to the human protein only.  
 CC The invention relates to 70 novel genes and their fragments (nucleic acid  
 CC sequences: V34154-V34276; amino acid sequences W75057-W75179) which  
 CC are useful for preventing, treating or ameliorating medical conditions  
 CC e.g. by protein or gene therapy. Also, pathological conditions can be  
 CC diagnosed by determining the amount of the new polypeptides in a sample  
 CC or by determining the presence of mutations in the new polynucleotides.  
 CC Specific uses are described for each of the 70 polynucleotides, based on  
 CC which tissues they are most highly expressed in (see V34154 for described  
 CC uses).  
 SQ Sequence 509 BP; 176 A; 122 C; 117 G; 94 T;

Query Match

2.2%; Score 18; DB 1; Length 509;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 809 catcccaaaaaaaaaaaaaa 826

|||||

Db 492 CATCCCAAAAAAAAAAAAAA 509

# RESULT 11

V44301

ID V44301 standard; cDNA; 572 BP.

AC V44301;

DT 06-OCT-1998 (first entry)

DE human secreted protein clone CP258\_3 cDNA.

KW Secreted protein; human adult salivary gland; nutrition; cytokine;

KW cell proliferation; differentiation; immune system; suppressor; ligand;

KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;

KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;

KW stimulant; anti-inflammatory; ds.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 40..456

FT /\*tag= a

FT /product= secreted protein

FT /note= "Isolated from clone CP258\_3"

PN WO9827205-A2.

PD 25-JUN-1998.

PF 17-DEC-1997; U23330.

PR 16-DEC-1997; US-991872.

PR 18-DEC-1996; US-769192.

PR 13-JAN-1997; US-783401.

PA (GEM ) GENETICS INST INC.

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,

PI Racine LA, Spaulding V, Treacy M;

DR WPI; 98-362774/31.

DR P-PSDB; W64226.

PT New polynucleotides and secreted proteins - obtained from human

PT foetal brain, human adult testes, human adult brain and human adult

PT salivary gland cDNA libraries

PS Claim 3a; Page 8; 10pp; English.

CC This sequence encodes a novel secreted protein from clone CP258\_3

CC isolated from a human adult salivary gland cDNA library. This protein

CC has applications for nutritional use, cytokine and cell

CC proliferation/differentiation activity, immune stimulating or

CC suppressing activity, hematopoiesis regulating activity, tissue growth

CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,

CC haemostatic and thrombotic activity, receptor/ligand activity,

CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,

CC tumour inhibition activity and other activities.

SQ Sequence 572 BP; 210 A; 132 C; 106 T;

# Query Match

Best Local Similarity 2.2%; Score 18; DB 1; Length 572;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 809 catcccaaaaaaaaaaaaaa 826

|||||

Db 531 CATCCCAAAAAAAAAAAAAA 548

# RESULT 12

V59522

ID V59522 standard; DNA; 653 BP.

AC V59522;

DT 06-JAN-1999 (first entry)

DE Human secreted protein gene 12 clone HCFW04.

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;

KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;

KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;

KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;

KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;

KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;

KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;

KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.

PN WO9839448-A2.

PD 11-SEP-1998.

PF 06-MAR-1998; U04493.

PR 02-OCT-1997; US-061060.

PR 07-MAR-1997; US-038621.

PR 07-MAR-1997; US-040161.

PR 07-MAR-1997; US-040162.

PR 07-MAR-1997; US-040163.

PR 07-MAR-1997; US-040333.

PR 07-MAR-1997; US-040334.

PR 07-MAR-1997; US-040336.

PR 11-APR-1997; US-040626.

PR 11-APR-1997; US-043311.

PR 11-APR-1997; US-043312.

PR 11-APR-1997; US-043313.

PR 11-APR-1997; US-043314.

PR 11-APR-1997; US-043568.

PR 11-APR-1997; US-043569.

PR 11-APR-1997; US-043576.

PR 11-APR-1997; US-043578.

PR 11-APR-1997; US-043580.

PR 11-APR-1997; US-043659.

PR 11-APR-1997; US-043670.

PR 11-APR-1997; US-043671.

PR 11-APR-1997; US-043672.

PR 11-APR-1997; US-043674.

PR 23-MAY-1997; US-047492.

PR 23-MAY-1997; US-047500.

PR 23-MAY-1997; US-047501.

PR 23-MAY-1997; US-047502.

PR 23-MAY-1997; US-047503.

PR 23-MAY-1997; US-047581.

PR 23-MAY-1997; US-047582.

PR 23-MAY-1997; US-047583.

PR 23-MAY-1997; US-047584.

PR 23-MAY-1997; US-047585.

PR 23-MAY-1997; US-047586.

PR 23-MAY-1997; US-047587.

PR 23-MAY-1997; US-047588.

PR 23-MAY-1997; US-047589.

PR 23-MAY-1997; US-047590.

PR 23-MAY-1997; US-047592.

PR 23-MAY-1997; US-047593.

PR 23-MAY-1997; US-047594.

PR 23-MAY-1997; US-047595.

PR 23-MAY-1997; US-047596.

PR 23-MAY-1997; US-047597.

PR 23-MAY-1997; US-047598.

PR 23-MAY-1997; US-047599.

PR 23-MAY-1997; US-047600.

PR 23-MAY-1997; US-047601.

PR 23-MAY-1997; US-047612.

PR 23-MAY-1997; US-047613.

PR 23-MAY-1997; US-047614.

PR 23-MAY-1997; US-047615.

PR 23-MAY-1997; US-047617.

PR 23-MAY-1997; US-047618.

PR 23-MAY-1997; US-047632.

PR 23-MAY-1997; US-047633.

PR 06-JUN-1997; US-048964.

PR 06-JUN-1997; US-048974.

PR 13-JUN-1997; US-049610.

PR 08-JUL-1997; US-051926.

PR 16-JUL-1997; US-052874.

PR 18-AUG-1997; US-055724.

PR 22-AUG-1997; US-056630.

PR 22-AUG-1997; US-056631.

PR 22-AUG-1997; US-056632.

PR 22-AUG-1997; US-056636.

PR 22-AUG-1997; US-056637.



AC	Q46775;
DE	12-JAN-1994 (first entry)
DT	Maize dwarf mosaic virus strain A (MDMV-A) coat protein gene.
KW	MDMV; coat protein; potyviruses; Zea mays; ss.
OS	Maize dwarf mosaic virus.
FH	Key Location/Qualifiers
FT	cds 1..876
FT	/tag= a
FT	/product= MDMV-A coat protein.
PN	WO9314210-A.
PD	22-JUL-1993.
PF	28-DEC-1992; EP03001.
PR	08-JAN-1992; US-817922.
PA	(SANO ) SANDOZ LTD.
PA	(SANO ) SANDOZ PATENT GMBH.
PA	(SANO ) SANDOZ-ERFINDUNGEN VERW GES MBH.
PI	Clark JM, Jilka JW, Murry LE, Scarafia LEC;
DR	WPI: 93-243225/30.
DR	P-PSDB; R39297.
PT	transgenic corn plant cells - comprise e.g. coat protein gene of
PT	maize dwarf mosaic viral strain, produces plant resistant to
PT	viral challenge
PS	Disclosure; Page 27-28; 69pp; English.
CC	The maize dwarf mosaic virus coat protein gene is used to transform
CC	corn plant cells. These are then used to produce plants resistant
CC	to a variety of viruses, particularly potyviruses and especially
CC	maize dwarf mosaic virus (MDMV). Other viral sequences which may be
CC	used with similar effect are antisense sequences (transcription
CC	inhibitors), helicases, defective interfering sequences, intergenic
CC	regions, 3' untranslated terminal sequences, viral helper components,
CC	viral movement protein genes, viral satellite genes, stem-loop
CC	sequences and transcriptional or translational factors of viral
CC	origin.
SQ	Sequence 1113 BP; 365 A; 210 C; 265 G; 273 U;

Query Match	2.2%	Score 18;	DB 1;	Length 1113;
Best Local Similarity	72.2%	Pred. No. 37;		
Matches 13;	Conservative 5;	Mismatches 0;	Indels 0;	Gaps 0;
QY	288	ctggtgatgttgatcca	305	
		: : : : : : : :		
Db	134	CUGGUGAUGGUGGAUCCA	151	
RESULT	14			
Q78278				
ID	Q78278	standard;	cdNA;	1348 BP.
AC	Q78278;			
DE	06-JUN-1995	(first entry)		
DT	T. niveum	GAPDH.		
KW	Alanine-racemase; glyceraldhyde-3-phosphate-dehydrogenase;			
KW	GAPDH; cyclosporin; immunosuppressive; ss.			
OS	Toypocladium niveum	ATCC 34921.		
PN	W09425606-A.			
PD	10-NOV-1994.			
PF	23-APR-1994;	E01272.		
PR	23-APR-1993;	DE-312856.		
PR	04-MAY-1993;	DE-314610.		

(SANOZ) SANDOZ WILD.  
 (SANOZ) SANDOZ PATENT GMBH.  
 (SANOZ) SANDOZ-ERFINDUNGEN VERW GES MBH.  
 KOCHER HP, SCHNEIDER-SCHERZER E, SCHOERGENDORFER K;  
 WEBER G;  
 WPI: 94-358281/44.  
 New nucleic acid encoding eukaryotic alanine racemase - and related vectors, host cells and recombinant enzyme, useful for producing cyclosporin derivs. or increasing cyclosporin prodn., also new glyceraldehyde-3-phosphate dehydrogenase gene.

PI also new glyceraldehyde-3-phosphate dehydrogenase gene.

PS Disclosure; Fig.12; 82pp; German.

CC A cosmid bank of *T. niveum* ATCC 34921 genomic DNA was screened with

CC a fragment of the glyceraldehyde-3-phosphate-dehydrogenase (GAPDH)

CC gene from Penicillium chrysogenum. A 2.2 kb fragment was  
 CC identified (Q78277) and cloned in pUC18 to give pGTL. A 665 bp  
 CC HindIII-HindII fragment of this was used to screen cosmid DNA to  
 CC identify a DNA fragment (Q78279) containing the GAPDH promoter.  
 CC A sequence (Q78278) encoding the GAPDH gene (gpda) was isolated from  
 CC a cDNA library. The full sequence of the genomic gpda structural  
 CC gene is given in Q78280 and the corresponding protein sequence in  
 CC R65965. The GAPDH promoter is useful for expressing foreign genes  
 CC in T. niveum.  
 SQ Sequence 1348 BP; 328 A; 448 C; 330 G; 242 T;

Query Match 2.2%; Score 18; DB 1; Length 1348;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 809 catcaaaaaaaaaaaaa 826  
 |||||  
 DB 1297 CATCCAAAAAAAAAAAA 1314

# RESULT 15

N80457  
 ID N80457 standard; cDNA; 1355 BP.  
 AC N80457;  
 DT 21-MAR-1991 (first entry)  
 DE Sequence of cDNA encoding 3' end of phenylalanine ammonia lyase  
 DE (PAL) derived from Rhodospiridium toruloides.  
 KW Phenylketonuria therapy; L-phenylalanine therapy; enzyme; ss.  
 OS Rhodospiridium toruloides.  
 FH Key  
 FT Location/Qualifiers  
 FT cds 22..1143  
 FT /\*tag= a  
 PN WO802024-A.  
 PD 24-MAR-1988.  
 PF 08-SEP-1987; G00628.  
 PR 08-SEP-1986; GB-021626.  
 PA (PUBL-) Public Health Lab (ANSO/).  
 PI Anson J, Gilbert H, Oram J, Minton NP;  
 DR WPI; 88-091814/13.  
 DR P-PSDB; P80277.  
 PT Gene encoding phenylalanine ammonia lyase - obtd. from a strain of  
 PT R. toruloides by excising non-coding introns.  
 PS Example; Fig 5; 43pp; English..  
 CC Figure 3 (N80456) shows a portion of the genetic DNA sequence of R.  
 CC toruloides which encodes PAL. The gene of the invention preferably  
 CC consists of a DNA sequence identical to, related to, derived from or  
 CC complementary to the sequence of codons from the start codon to the  
 CC stop codon in Fig 3, from which the six intron IVS1 to IVS6 have  
 CC been deleted (see N82309). The gene from which introns have been  
 CC deleted is prepd. by joining overlapping cDNA clones pPAL1 (Fig 5,  
 CC N80457) and pPAL2 (Fig 6, N80458). As the introns present in the  
 CC natural PAL gene act as a barrier to the expression of PAL in  
 CC organisms other than R. toruloides, the gene can be used to produce  
 CC PAL in a wide range of procaryotic and eukaryotic hosts.  
 SQ Sequence 1355 BP; 273 A; 521 C; 316 G; 245 T;

Query Match 2.2%; Score 18; DB 1; Length 1355;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 809 catcaaaaaaaaaaaaa 826  
 |||||  
 DB 1304 CATCCAAAAAAAAAAAA 1321

Search completed: May 1, 2000, 18:54:50  
 Job time: 18974 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 16:11:14 ; Search time 181.17 Seconds  
(without alignments)  
545.848 Million cell updates/sec

Title: US-09-215-435-124

Perfect score: 826  
Sequence: 1 aacacagggtgccatgggt.....atcatccaaaaa 826

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 214294 seqs, 59861574 residues

Word size : 0

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PCRU9\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	2.2	1113	1 US-08-229-287-1	Sequence 1, Appl
2	18	2.2	1485	4 US-08-861-512-2	Sequence 2, Appl
3	18	2.2	2783	1 US-08-191-337-1	Sequence 1, Appl
4	17	2.1	260	7 5514558-9	Patent No. 5514558
5	17	2.1	516	2 US-08-288-888-1	Sequence 1, Appl
6	17	2.1	517	3 US-08-910-075-1	Sequence 1, Appl
7	17	2.1	517	4 US-08-905-801A-1	Sequence 1, Appl
8	17	2.1	581	1 US-07-620-426B-30	Sequence 30, Appl
9	17	2.1	581	1 US-07-662-007B-36	Sequence 36, Appl
10	17	2.1	581	1 US-07-824-247-36	Sequence 36, Appl
11	17	2.1	898	1 US-08-419-078-3	Sequence 3, Appl
12	17	2.1	898	1 US-08-726-883-3	Sequence 3, Appl
13	17	2.1	987	4 US-08-891-837B-1	Sequence 1, Appl
14	17	2.1	1064	3 US-08-378-588-15	Sequence 15, Appl
15	17	2.1	1064	3 US-08-811-094-15	Sequence 15, Appl
16	17	2.1	1064	6 PCR-US94-11121-15	Sequence 15, Appl
17	17	2.1	1208	2 US-07-949-516A-3	Sequence 3, Appl
18	17	2.1	1208	3 US-08-814-459-3	Sequence 3, Appl
19	17	2.1	1270	3 US-08-378-588-23	Sequence 23, Appl
20	17	2.1	1270	3 US-08-811-094-23	Sequence 23, Appl
21	17	2.1	1270	6 PCR-US94-11121-23	Sequence 23, Appl
22	17	2.1	1371	1 US-08-181-271A-43	Sequence 43, Appl
23	17	2.1	1371	1 US-08-449-315-43	Sequence 43, Appl
24	17	2.1	1371	1 US-08-444-803-43	Sequence 43, Appl
25	17	2.1	1371	1 US-08-449-043-43	Sequence 43, Appl
26	17	2.1	1371	2 US-08-456-265A-43	Sequence 43, Appl

27 17 2.1 1371 2 US-08-455-416-43 Sequence 43, Appl  
28 17 2.1 1371 2 US-08-455-244-43 Sequence 43, Appl  
29 17 2.1 1371 2 US-08-454-876-43 Sequence 43, Appl  
30 17 2.1 1371 3 US-08-457-364-43 Sequence 43, Appl  
31 17 2.1 1371 3 US-08-456-262-43 Sequence 43, Appl  
32 17 2.1 1371 3 US-08-456-240-43 Sequence 43, Appl  
33 17 2.1 1371 3 US-08-455-736-43 Sequence 43, Appl  
34 17 2.1 1371 4 US-08-971-217-43 Sequence 43, Appl  
35 17 2.1 1389 2 US-08-702-344-20 Sequence 20, Appl  
36 17 2.1 1525 1 US-08-186-833-3 Sequence 3, Appl  
37 17 2.1 1525 2 US-08-609-572-1 Sequence 1, Appl  
38 17 2.1 1591 1 US-08-728-956-3 Sequence 3, Appl  
39 17 2.1 1719 2 US-08-472-028A-1 Sequence 1, Appl  
40 17 2.1 1719 4 US-08-808-931-1 Sequence 1, Appl  
41 17 2.1 1813 3 US-08-007-107-1 Sequence 1, Appl  
42 17 2.1 2103 4 US-08-897-340-2 Sequence 2, Appl  
43 17 2.1 2277 2 US-08-676-967-5 Sequence 5, Appl  
44 17 2.1 2277 2 US-08-676-974-5 Sequence 5, Appl  
45 17 2.1 2277 4 US-09-098-487-5 Sequence 5, Appl

#### ALIGNMENTS

RESULT 1  
US-08-229-287-1  
; Sequence 1, Application US/08229287  
; Patent No. 5530193  
; GENERAL INFORMATION:  
; APPLICANT: Clark Jr., John M.  
; APPLICANT: Jilka, Joseph M.  
; APPLICANT: Murry, Lynn E.  
; APPLICANT: Scarafia, Lilliana E.  
; TITLE OF INVENTION: VIRUS RESISTANT CORN PLANTS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sandoz Agro, Inc.  
; STREET: 975 California Avenue  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/229,287  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/067,257  
; FILING DATE: 25-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/817,922  
; FILING DATE: 08-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Marcus-Wyner, Lynn  
; REGISTRATION NUMBER: 34,869  
; REFERENCE/DOCKET NUMBER: 135-1084/XCC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/354-3588  
; TELEFAX: 415/857-1125  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1113 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..876



```
Db 77 AAAAAAAAAATCATTCCA 61
|||||
; TITLE OF INVENTION: ALLERGEN, HOST CELLS CONTAINING SUCH DNA AND METHODS
; TITLE OF INVENTION: FOR PRODUCING Derf II (as amended)
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,075
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-50848
; FILING DATE: 03-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/658,596
; FILING DATE: 21-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/288,888
; FILING DATE: 10-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: YUUKI-ID
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..426
; US-08-910-075-1

Query Match 2.1%; Score 17; DB 3; Length 517;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 810 atcccaaaaaaaaaa 826
|||||
Db 490 ATCCAAAAAAAAAAAA 506

RESULT 7
US-08-905-801A-1
; Sequence 1, Application US/08905801A
; Patent No. 5958415
; GENERAL INFORMATION:
; APPLICANT: YUUKI, Toshifumi
; APPLICANT: OKUMURA, Yasushi
; APPLICANT: YAMAKAWA, Hiroshi
; TITLE OF INVENTION: A PROTEIN ALLERGEN OF Derf II AND
; TITLE OF INVENTION: COMPOSITION (as amended)
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.

Db 77 AAAAAAAAAATCATTCCA 61
|||||
; TITLE OF INVENTION: ALLERGEN, HOST CELLS CONTAINING SUCH DNA AND METHODS
; TITLE OF INVENTION: FOR PRODUCING Derf II (as amended)
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/288,888
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/658,596
; FILING DATE: 21-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Cooper, Iver P
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: YUUK11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 516 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..426
; US-08-288-888-1

Query Match 2.1%; Score 17; DB 2; Length 516;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 810 atcccaaaaaaaaaa 826
|||||
Db 490 ATCCAAAAAAAAAAAA 506

RESULT 6
US-08-910-075-1
; Sequence 1, Application US/08910075
; Patent No. 5876722
; GENERAL INFORMATION:
; APPLICANT: YUUKI, Toshifumi
; APPLICANT: OKUMURA, Yasushi
; APPLICANT: YAMAKAWA, Hiroshi
; TITLE OF INVENTION: DNA ENCODING Derf II, THE MAJOR MITE
```

```
;
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,801A
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-50848
; FILING DATE: 03-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/658,596
; FILING DATE: 21-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/288,888
; FILING DATE: 10-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: YUUKI-1C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..426
;
US-08-905-801A-1

Query Match 2.1%; Score 17; DB 4; Length 517;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 810 atccaaaaaa 826
Db 490 ATCCAAAAAA 506

RESULT 8
US-07-620-426B-30
; Sequence 30, Application US/07620426B
; Patent No. 5298421
; GENERAL INFORMATION:
; APPLICANT: Davies, Huw Maelor
; APPLICANT: Pollard, Michael Roman
; APPLICANT: Voekler, Toni Alois
; TITLE OF INVENTION: Plant Thioesterases
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: California
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 KB storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.7
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/620,426B
; FILING DATE: 19910408
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/514,030
; FILING DATE: 26-APR-1990
; APPLICATION NUMBER: 07/620,426
; FILING DATE: 30-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: CGNE 70-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 916-753-6313
; TELEFAX: 916-753-1510
; INFORMATION FOR SEQ ID NO: 36:
```

```
;
; FILING DATE: 19901130
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/514,030
; FILING DATE: 26-APRIL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; REFERENCE/DOCKET NUMBER: CGNE 70-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 916-753-6313
; TELEFAX: 916-753-1510
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 581 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
;
US-07-620-426B-30

Query Match 2.1%; Score 17; DB 1; Length 581;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 810 atccaaaaaa 826
Db 564 ATCCAAAAAA 580

RESULT 9
US-07-662-007B-36
; Sequence 36, Application US/07662007B
; Patent No. 534771
; GENERAL INFORMATION:
; APPLICANT: Davies, Huw Maelor
; APPLICANT: Pollard, Michael Roman
; APPLICANT: Voekler, Toni Alois
; APPLICANT: Thompson, Gregory A.
; TITLE OF INVENTION: Plant Thioesterases
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: California
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 KB storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.7
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/662,007B
; FILING DATE: 19910408
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/514,030
; FILING DATE: 26-APR-1990
; APPLICATION NUMBER: 07/620,426
; FILING DATE: 30-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: CGNE 70-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 916-753-6313
; TELEFAX: 916-753-1510
; INFORMATION FOR SEQ ID NO: 36:
```

```

Query Match      2.1%; Score 17; DB 1; Length 898;
Best Local Similarity 100.0%; Pred. NO. 56;
Matches 17; Conservative 0; Mismatches 0; Indels
y      810 atccaaaaaaaaaaaa 826
      |||||
b      877 ATCCAAAAAANAANA 893

RESULT 12

```

```
US-08-726-883-3
; Sequence 3, Application US/08726883
; Patent No. 5676946
; GENERAL INFORMATION:
; APPLICANT: HAWKINS, PHILLIP R.
; APPLICANT: SELINGER, JEFFREY J.
; TITLE OF INVENTION: PHOSPHOLIPASE C HOMOLOG
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3330 HILLVIEW AVENUE
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,883
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/419,078
; FILING DATE: 10-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: LUTHER, BARBARA J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF0030 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0355
; TELEFAX: 415-855-0572
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 898 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: Human Mast Cell
; CLONE: 9118
; US-08-726-883-3

Query Match          2.1%; Score 17; DB 1; Length 898;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 810 atccaaataaaataaa 826
|||||
Db 877 ATCCAAAAAATAAA 893

RESULT 13
US-08-891-837B-1
; Sequence 1, Application US/08891837B
; Patent No. 5968771
; GENERAL INFORMATION:
; APPLICANT: Jin, Kun Lin
; APPLICANT: Chen, Jun
; APPLICANT: Graham, Steven H.
; APPLICANT: Simon, Roger P.
; TITLE OF INVENTION: Global Ischemia Induced Gene
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: University of Pittsburgh
; STREET: Office of Technology Transfer
; STREET: 911 William Pitt Union
; CITY: Pittsburgh
; STATE: Pennsylvania

Query Match          2.1%; Score 17; DB 1; Length 898;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 810 atccaaataaaataaa 826
|||||
Db 877 ATCCAAAAAATAAA 893

RESULT 14
US-08-378-588-15
; Sequence 15, Application US/08378588
; Patent No. 5608148
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: TRANSGENIC COTTON PLANTS
; TITLE OF INVENTION: PRODUCING HETEROLOGOUS PEROXIDASE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: First Wisconsin Plaza, One South
; STREET: Pinckney St.,
; STREET: P.O. Box 2113
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/378,588
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 11-229-9101-2
; TELECOMMUNICATION INFORMATION:
```



; TELEPHONE: (608) 251-2484  
; TELEFAX: (608) 251-9166  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1064 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; US-08-378-588-15

Query Match 2.1%; Score 17; DB 1; Length 1064;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 810 atccaaaaaa 826  
|||||  
Db 1029 ATCCAAAAAA 1045

## RESULT 15

US-08-811-094-15  
; Sequence 15, Application US/08811094  
; Patent No. 5869720

; GENERAL INFORMATION:  
; APPLICANT: John, Maliyakal E.  
; TITLE OF INVENTION: TRANSGENIC COTTON PLANTS  
; TITLE OF INVENTION: PRODUCING HETEROLOGOUS PEROXIDASE  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nicholas J. Seay, Quarles & Brady  
; STREET: First Wisconsin Plaza, One South  
; STREET: Pinckney St.,  
; STREET: P.O. Box 2113  
; CITY: Madison  
; STATE: WI  
; COUNTRY: USA  
; ZIP: 53701-2113

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811,094  
; FILING DATE: 03-MAR-1997

## CLASSIFICATION: 800

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/378,588

; FILING DATE: 25-JAN-1995  
; ATTORNEY/AGENT INFORMATION:

; NAME: Seay, Nicholas J.

; REGISTRATION NUMBER: 27,386

; REFERENCE/DOCKET NUMBER: 11-229-9101-2

## TELECOMMUNICATION INFORMATION:

; TELEPHONE: (608) 251-2484

; TELEFAX: (608) 251-9166

; INFORMATION FOR SEQ ID NO: 15:

## SEQUENCE CHARACTERISTICS:

; LENGTH: 1064 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cdna

; US-08-811-094-15

Query Match 2.1%; Score 17; DB 3; Length 1064;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 810 atccaaaaaa 826

Db 1029 ATCCAAAAAA 1045  
|||||

Search completed: May 1, 2000, 16:11:19  
Job time: 18744 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 20:00:21 ; Search time 4088.29 Seconds  
(without alignments)  
-692.841 Million cell updates/sec

Title: US-09-215-435-124  
Perfect score: 826  
Sequence: 1 aacagaggtgccatgggt.....atcatccaaaaa

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 5265378 seqs, -1714608768 residues

Word size : 0

Total number of hits satisfying chosen parameters: 10530736

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Pending Patents NA.\*  
1: /cgn2\_6/ptodata/1/pna/PCTUS9\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/pna/US081A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US081B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US081C\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US081D\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US082A\_COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US082B\_COMB.seq.\*  
9: /cgn2\_6/ptodata/1/pna/US082C\_COMB.seq.\*  
10: /cgn2\_6/ptodata/1/pna/US083A\_COMB.seq.\*  
11: /cgn2\_6/ptodata/1/pna/US083B\_COMB.seq.\*  
12: /cgn2\_6/ptodata/1/pna/US084A\_COMB.seq.\*  
13: /cgn2\_6/ptodata/1/pna/US084B\_COMB.seq.\*  
14: /cgn2\_6/ptodata/1/pna/US084C\_COMB.seq.\*  
15: /cgn2\_6/ptodata/1/pna/US084D\_COMB.seq.\*  
16: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*  
17: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*  
18: /cgn2\_6/ptodata/1/pna/US087A\_COMB.seq.\*  
19: /cgn2\_6/ptodata/1/pna/US087B\_COMB.seq.\*  
20: /cgn2\_6/ptodata/1/pna/US087C\_COMB.seq.\*  
21: /cgn2\_6/ptodata/1/pna/US088A\_COMB.seq.\*  
22: /cgn2\_6/ptodata/1/pna/US088B\_COMB.seq.\*  
23: /cgn2\_6/ptodata/1/pna/US088C\_COMB.seq.\*  
24: /cgn2\_6/ptodata/1/pna/US088D\_COMB.seq.\*  
25: /cgn2\_6/ptodata/1/pna/US089A\_COMB.seq.\*  
26: /cgn2\_6/ptodata/1/pna/US089B\_COMB.seq.\*  
27: /cgn2\_6/ptodata/1/pna/US089C\_COMB.seq.\*  
28: /cgn2\_6/ptodata/1/pna/US089D\_COMB.seq.\*  
29: /cgn2\_6/ptodata/1/pna/US089E\_COMB.seq.\*  
30: /cgn2\_6/ptodata/1/pna/US089F\_COMB.seq.\*  
31: /cgn2\_6/ptodata/1/pna/US089G\_COMB.seq.\*  
32: /cgn2\_6/ptodata/1/pna/US089H\_COMB.seq.\*  
33: /cgn2\_6/ptodata/1/pna/US089I\_COMB.seq.\*  
34: /cgn2\_6/ptodata/1/pna/US090A\_COMB.seq.\*  
35: /cgn2\_6/ptodata/1/pna/US090B\_COMB.seq.\*  
36: /cgn2\_6/ptodata/1/pna/US090C\_COMB.seq.\*  
37: /cgn2\_6/ptodata/1/pna/US090D\_COMB.seq.\*  
38: /cgn2\_6/ptodata/1/pna/US090E\_COMB.seq.\*  
39: /cgn2\_6/ptodata/1/pna/US091A\_COMB.seq.\*  
40: /cgn2\_6/ptodata/1/pna/US091B\_COMB.seq.\*  
41: /cgn2\_6/ptodata/1/pna/US091C\_COMB.seq.\*  
42: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq.\*  
43: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq.\*

44: /cgn2\_6/ptodata/1/pna/US092C\_COMB.seq.\*  
45: /cgn2\_6/ptodata/1/pna/US092D\_COMB.seq.\*  
46: /cgn2\_6/ptodata/1/pna/US092E\_COMB.seq.\*  
47: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq.\*  
48: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq.\*  
49: /cgn2\_6/ptodata/1/pna/US093C\_COMB.seq.\*  
50: /cgn2\_6/ptodata/1/pna/US093D\_COMB.seq.\*  
51: /cgn2\_6/ptodata/1/pna/US093E\_COMB.seq.\*  
52: /cgn2\_6/ptodata/1/pna/US093F\_COMB.seq.\*  
53: /cgn2\_6/ptodata/1/pna/US094A\_COMB.seq.\*  
54: /cgn2\_6/ptodata/1/pna/US094B\_COMB.seq.\*  
55: /cgn2\_6/ptodata/1/pna/US094C\_COMB.seq.\*  
56: /cgn2\_6/ptodata/1/pna/US094D\_COMB.seq.\*  
57: /cgn2\_6/ptodata/1/pna/US094E\_COMB.seq.\*  
58: /cgn2\_6/ptodata/1/pna/US094F\_COMB.seq.\*  
59: /cgn2\_6/ptodata/1/pna/US094G\_COMB.seq.\*  
60: /cgn2\_6/ptodata/1/pna/US094H\_COMB.seq.\*  
61: /cgn2\_6/ptodata/1/pna/US094I\_COMB.seq.\*  
62: /cgn2\_6/ptodata/1/pna/US094J\_COMB.seq.\*  
63: /cgn2\_6/ptodata/1/pna/US094K\_COMB.seq.\*  
64: /cgn2\_6/ptodata/1/pna/US094L\_COMB.seq.\*  
65: /cgn2\_6/ptodata/1/pna/US094M\_COMB.seq.\*  
66: /cgn2\_6/ptodata/1/pna/US094N\_COMB.seq.\*  
67: /cgn2\_6/ptodata/1/pna/US094O\_COMB.seq.\*  
68: /cgn2\_6/ptodata/1/pna/US094P\_COMB.seq.\*  
69: /cgn2\_6/ptodata/1/pna/US094Q\_COMB.seq.\*  
70: /cgn2\_6/ptodata/1/pna/US094R\_COMB.seq.\*  
71: /cgn2\_6/ptodata/1/pna/US094S\_COMB.seq.\*  
72: /cgn2\_6/ptodata/1/pna/US094T\_COMB.seq.\*  
73: /cgn2\_6/ptodata/1/pna/US094U\_COMB.seq.\*  
74: /cgn2\_6/ptodata/1/pna/US094V\_COMB.seq.\*  
75: /cgn2\_6/ptodata/1/pna/US094W\_COMB.seq.\*  
76: /cgn2\_6/ptodata/1/pna/US094X\_COMB.seq.\*  
77: /cgn2\_6/ptodata/1/pna/US094Y\_COMB.seq.\*  
78: /cgn2\_6/ptodata/1/pna/US094Z\_COMB.seq.\*  
79: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq.\*  
80: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq.\*  
81: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq.\*  
82: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq.\*  
83: /cgn2\_6/ptodata/1/pna/US095E\_COMB.seq.\*  
84: /cgn2\_6/ptodata/1/pna/US095F\_COMB.seq.\*  
85: /cgn2\_6/ptodata/1/pna/US095G\_COMB.seq.\*  
86: /cgn2\_6/ptodata/1/pna/US095H\_COMB.seq.\*  
87: /cgn2\_6/ptodata/1/pna/US095I\_COMB.seq.\*  
88: /cgn2\_6/ptodata/1/pna/US095J\_COMB.seq.\*  
89: /cgn2\_6/ptodata/1/pna/US095K\_COMB.seq.\*  
90: /cgn2\_6/ptodata/1/pna/US095L\_COMB.seq.\*  
91: /cgn2\_6/ptodata/1/pna/US095M\_COMB.seq.\*  
92: /cgn2\_6/ptodata/1/pna/US095N\_COMB.seq.\*  
93: /cgn2\_6/ptodata/1/pna/US095O\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	826	100.0	826	42	US-09-215-435-124
2	826	100.0	826	54	US-09-471-276-7
3	826	100.0	826	69	US-60-081-563-70
4	826	100.0	826	84	US-60-147-499-7
5	826	100.0	826	86	US-60-169-629-7
6	826	100.0	826	93	US-60-187-470-7
7	610	73.8	1028	87	US-60-172-373-10894
8	610	73.8	1028	1	PCT-US99-29950-56
9	559	67.7	878	50	US-09-359-922-5157
10	559	67.7	878	92	US-09-359-922-5157
11	559	67.7	903	29	US-08-958-820-4
12	559	67.7	903	42	US-09-208-718-4
13	559	67.7	903	92	US-09-390-126-4
					Sequence 124, Appl
					Sequence 7, Appl
					Sequence 70, Appl
					Sequence 7, Appl
					Sequence 7, Appl
					Sequence 10894, A
					Sequence 56, Appl
					Sequence 5157, Ap
					Sequence 5157, Ap
					Sequence 4, Appl
					Sequence 4, Appl

```
14 559 67.7 903 92 US-09-390-126-4
15 479 58.0 952 1 PCT-US99-29950-61
16 335 40.6 886 92 US-09-491-404-1525
17 320 38.7 430 46 US-09-293-972-28063
18 319 38.6 394 50 US-09-359-067-3691
19 281 34.0 356 50 US-09-359-067-4300
20 277 33.5 353 50 US-09-359-067-4523
21 275 33.3 447 25 US-08-905-144-182
22 264 32.0 264 25 US-08-901-904-1660
23 264 32.0 408 50 US-09-359-067-2207
24 256 31.0 256 25 US-08-901-904-870
25 253 28.2 476 50 US-09-359-067-2655
26 226 27.4 429 48 US-09-332-782-5859
27 225 27.2 251 39 US-09-105-427-377
28 213 25.8 264 21 US-08-808-904-4221
29 213 25.8 314 21 US-08-808-904-4221
30 206 24.9 314 21 US-08-808-904-4221
31 206 24.9 314 21 US-08-808-904-4221
32 206 24.9 314 21 US-08-808-904-4221
33 194 23.5 273 24 US-08-880-314-4295
34 192 23.2 301 37 US-09-062-736-294
35 192 23.2 301 64 US-09-045-650-294
36 188 22.8 239 35 US-09-035-171-1718
37 188 22.8 239 62 US-09-035-171-1718
38 188 22.8 241 23 US-08-853-295-177
39 188 22.8 241 57 US-06-017-601-177
40 176 21.3 229 23 US-08-853-295-180
41 176 21.3 229 57 US-06-017-601-180
42 169 20.5 271 28 US-08-941-479-1345
43 160 19.4 266 44 US-09-271-490-17027
44 157 19.0 263 44 US-09-271-490-17026
45 151 18.3 220 24 US-08-899-858-1668
```

## ALIGNMENTS

```
RESULT 1
US-09-215-435-124
; Sequence 124, Application US/09215435A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert Aymeric,
; TITLE OF INVENTION: Extended CDNAS
; FILE REFERENCE: GENSET.019A
; CURRENT APPLICATION NUMBER: US/09/215,435A
; EARLIER FILING DATE: 1998-12-17
; EARLIER FILING DATE: 1997-12-17
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-2-9
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-4-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-8-10
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: Patent.pm
; SEQ ID NO 124
; LENGTH: 826
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 15..695
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 15..80
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: seq 8.5
; OTHER INFORMATION: seq AALLGLMVMVTG/DE
; FEATURE:
; NAME/KEY: polyA_signal
```

```
; LOCATION: 795..800
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 814..826
US-09-215-435-124

Query Match      100.0%; Score 826; DB 42; Length 826;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 826; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aaccagagtgccctggttggaacatgagctggtcacagcagcactgttactgggtc 60
   |||
Db 1 aaccagagtgccctggttggaacatgagctggtcacagcagcactgttactgggtc 60
   |||

QY 61 tcattgatggtgctactgagacgagagatgaaacagcccgtgtcccatgagggccctc 120
   |||
Db 61 tcattgatggtgctactgagacgagagatgaaacagcccgtgtgtcccatgagggccctc 120
   |||

QY 121 tggcagagagacaccctcttttggcaggggccttgaagtttttaccacagagttggggaaca 180
   |||
Db 121 tggcagagagacaccctcttttggcaggggccttgaagtttttaccacagagttggggaaca 180
   |||

QY 181 ttggctgcaagttgttctctgatttaacactacacagagagatcactcctctgattg 240
   |||
Db 181 ttggctgcaagttgttctctgatttaacactacacagagagatcactcctctgattg 240
   |||

QY 241 agccgatagtcgaagttcccgggggcgtggacggcgcaacctatatctctgggtgattggtg 300
   |||
Db 241 agccgatagtcgaagttcccgggggcgtggacggcgcaacctatatctctgggtgattggtg 300
   |||

QY 301 atccgatgcccctagcagagacagaccacagacagagattcttggagacatttggctggttaa 360
   |||
Db 301 atccgatgcccctagcagagacagaccacagacagagattcttggagacatttggctggttaa 360
   |||

QY 361 cagatacaaggggccacacctgaagaaaggaaattcagggccacagagttatcagcct 420
   |||
Db 361 cagatacaaggggccacacctgaagaaaggaaattcagggccacagagttatcagcct 420
   |||

QY 421 accaggtccctcccccacggcgacacagtggttccatcgctaccagttctttgtctatc 480
   |||
Db 421 accaggtccctcccccacggcgacacagtggttccatcgctaccagttctttgtctatc 480
   |||

QY 481 ttcaggaagaaaggtcatctctctctcccaaggaaacacacacagagctcttcttga 540
   |||
Db 481 ttcaggaagaaaggtcatctctctctcccaaggaaacacacacagagctcttcttga 540
   |||

QY 541 aaatggacagatttctgaacgtttccacgtggcggaacctgaagcaagcaccagttca 600
   |||
Db 541 aaatggacagatttctgaacgtttccacgtggcggaacctgaagcaagcaccagttca 600
   |||

QY 601 tgaccacgaactaccagagactcaccaacctccaggctccacagagaaagggccagcgagc 660
   |||
Db 601 tgaccacgaactaccagagactcaccaacctccaggctccacagagaaagggccagcgagc 660
   |||

QY 661 ccaagcacaacacacagcgagatagctgctctagatagccgctttgcatcccgag 720
   |||
Db 661 ccaagcacaacacacagcgagatagctgctctagatagccgctttgcatcccgag 720
   |||

QY 721 catgtggccacactgcccaccacggacgagtggtggtatggaacccctctggtacagaa 780
   |||
Db 721 catgtggccacactgcccaccacggacgagtggtggtatggaacccctctggtacagaa 780
   |||

QY 781 cccctcttttccaaataaaaaaaatcatcccaaaaaaa 826
   |||
Db 781 cccctcttttccaaataaaaaaaatcatcccaaaaaaa 826
   |||
```

```
RESULT 2
US-09-471-276-7
; Sequence 7, Application US/09471276
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
```

```
; APPLICANT: Duclert A.
; APPLICANT: Giordano, J.Y.
; FILE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: GENSET.025CP1
; CURRENT APPLICATION NUMBER: US/09/471,276
; CURRENT FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: 09/057,719
; EARLIER FILING DATE: 1998-04-09
; EARLIER APPLICATION NUMBER: 09/069,047
; EARLIER FILING DATE: 1998-04-28
; EARLIER APPLICATION NUMBER: PCT/IB99/00712
; EARLIER FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 1622
; SOFTWARE: Patent.pm
; SEQ ID NO 7
; LENGTH: 826
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 15..695
; NAME/KEY: sig_peptide
; LOCATION: 15..80
; OTHER INFORMATION: score 8.5
; OTHER INFORMATION: seq AALLGLMMVVTG/DE
US-09-471-276-7

Query Match      100.0%; Score 826; DB 54; Length 826;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 826; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aaccagagtgccatgggttgacaaatgagctgtgtcacagcagcagcactgttactgggtc 60
Db 1 aaccagagtgccatgggttgacaaatgagctgtgtcacagcagcagcactgttactgggtc 60
QY 61 tcatgatgtgttcaactgagacagagatgagaaacacgcggtgtgcccattgagggccctcc 120
Db 61 tcatgatgtgttcaactgagacagagatgagaaacacgcggtgtgcccattgagggccctcc 120
QY 121 tggacagagacaccccttttgcagggccttgaaagtttctaccacagagttggggaaca 180
Db 121 tggacagagacaccccttttgcagggccttgaaagtttctaccacagagttggggaaca 180
QY 181 ttggctgcaaggtgttctctgattgttaacaactacagacagagatcacctcctggatgg 240
Db 181 ttggctgcaaggtgttctctgattgttaacaactacagacagagatcacctcctggatgg 240
QY 241 agccgatagtcgaagttcccgggggccggtgagcggcgcaacccctatatcctgggtggtgg 300
Db 241 agccgatagtcgaagttcccgggggccggtgagcggcgcaacccctatatcctgggtggtgg 300
QY 301 attcagatgccccttagcagacagacacccacagagattctggagacatttggctgggttaa 360
Db 301 attcagatgccccttagcagacagacacccacagagattctggagacatttggctgggttaa 360
QY 361 cagatatcaaggcggccgacccctgaagaagggaagattcaggggccagggttatcagcct 420
Db 361 cagatatcaaggcggccgacccctgaagaagggaagattcaggggccagggttatcagcct 420
QY 421 accaggtcctccctccacacgggacacagtggttccatcgctaccagttctttgtctatc 480
Db 421 accaggtcctccctccacacgggacacagtggttccatcgctaccagttctttgtctatc 480
QY 481 ttcaggaagaaaggtcatctctctcttcccaaggaagaaacaaactcagagctcttga 540
Db 481 ttcaggaagaaaggtcatctctctcttcccaaggaagaaacaaactcagagctcttga 540
QY 541 aaatggacagatttctgaacccgtttccacctggcggaacctgaagcaagcaccagttca 600
Db 541 aaatggacagatttctgaacccgtttccacctggcggaacctgaagcaagcaccagttca 600
QY 601 tgaccagaactaccagggactcccaacccctccaggtctccacagagagaaggccagcgagc 660
```

```
Db 601 tgaccagaactaccaggaactcaccacccctccaggtctccacagagaaggccagcgagc 660
QY 661 ccaagcacaaaaaacccagcgagatagctgctgttagatagccggtttgcatccggg 720
Db 661 ccaagcacaaaaaacccagcgagatagctgctgttagatagccggtttgcatccggg 720
QY 721 catgtggccacactgcccaccaccacacgatgtgggtgtatggaacccctcttgatacagaa 780
Db 721 catgtggccacactgcccaccaccacacgatgtgggtgtatggaacccctcttgatacagaa 780
QY 781 cccctcttttccaaataaaaaaaatcatccaaaaaataaaaaa 826
Db 781 cccctcttttccaaataaaaaaaatcatccaaaaaataaaaaa 826

RESULT 3
US-60-081-563-70
; Sequence 70, Application US/60081563
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste Dumas Milne
; APPLICANT: Duclert, Aymeric
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SECRETED PROTEINS
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/081,563
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.027PR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Prostate
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 15..695
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 15..80
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 8.5
; OTHER INFORMATION: seq AALLGLMMVVTG/DE
; FEATURE:
; NAME/KEY: poly_a_signal
; LOCATION: 795..800
; FEATURE:
; NAME/KEY: poly_a
; LOCATION: 814..826
; IDENTIFICATION METHOD: blastn
```

```
FEATURE:
NAME/KEY: est
LOCATION: 426..664
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 99..337
OTHER INFORMATION: id W37255
FEATURE:
NAME/KEY: est
LOCATION: 327..435
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 1..109
OTHER INFORMATION: id W37255
FEATURE:
NAME/KEY: est
LOCATION: 716..777
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 391..452
OTHER INFORMATION: id W37255
FEATURE:
NAME/KEY: est
LOCATION: 660..720
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 334..394
OTHER INFORMATION: id W37255
FEATURE:
NAME/KEY: est
LOCATION: 506..630
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 90
OTHER INFORMATION: region 327..451
OTHER INFORMATION: id W32197
FEATURE:
NAME/KEY: est
LOCATION: 323..435
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 94
OTHER INFORMATION: region 143..255
OTHER INFORMATION: id W32197
FEATURE:
NAME/KEY: est
LOCATION: 180..280
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 98
OTHER INFORMATION: region 2..102
OTHER INFORMATION: id W32197
FEATURE:
NAME/KEY: est
LOCATION: 464..528
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 284..348
OTHER INFORMATION: id W32197
FEATURE:
NAME/KEY: est
LOCATION: 279..321
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 100..142
OTHER INFORMATION: id W32197
FEATURE:
NAME/KEY: est
LOCATION: 426..466
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 97
OTHER INFORMATION: region 245..285
OTHER INFORMATION: id W32197
FEATURE:
NAME/KEY: est
LOCATION: 358..674
```

```
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 98
OTHER INFORMATION: region 141..457
OTHER INFORMATION: id AAL92427
FEATURE:
NAME/KEY: est
LOCATION: 670..779
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 37..146
OTHER INFORMATION: id AAL92427
FEATURE:
NAME/KEY: est
LOCATION: 485..602
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 224..341
OTHER INFORMATION: id W33189
FEATURE:
NAME/KEY: est
LOCATION: 601..709
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 91
OTHER INFORMATION: region 115..223
OTHER INFORMATION: id W33189
FEATURE:
NAME/KEY: est
LOCATION: 704..779
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 98
OTHER INFORMATION: region 46..121
OTHER INFORMATION: id W33189
FEATURE:
NAME/KEY: est
LOCATION: 392..436
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 97
OTHER INFORMATION: region 393..437
OTHER INFORMATION: id W33189
FEATURE:
NAME/KEY: est
LOCATION: 429..465
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 94
OTHER INFORMATION: region 363..399
OTHER INFORMATION: id W33189
FEATURE:
NAME/KEY: est
LOCATION: 464..493
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 334..363
OTHER INFORMATION: id W33189
FEATURE:
NAME/KEY: est
LOCATION: 343..730
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 98
OTHER INFORMATION: region 8..395
OTHER INFORMATION: id HSPD04880
FEATURE:
NAME/KEY: est
LOCATION: 507..663
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 264..420
OTHER INFORMATION: id AAL92426
US-60-081-563-70
```

```
Query Match 100.0% Score 826; DB 69; Length 826;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 826; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy 1 aaccagaggtgccatgggttggacaatgagctggttcacagcagcactgttactgggtc 60
Db 1 AACCAAGGTGCCATGGGTGGACAATGAGCTGGTTCACAGCAGCAGCTTACTGGGTC 60

Qy 61 tcatgatggttgcactgagacagaggtgagaacagcccgctgtgccatgagccctcc 120
Db 61 TCATGATGGTGTCACTGGAGCAGGAGTGAACAGCCCGTGTGCCATGAGGCCCTCC 120

Qy 121 tggcagagacacacctcttttggcagggccttgaagttttaccagagttggggaaca 180
Db 121 TGGCAGGAGACACCTCTTTGGCAGGGCCTGAAGTTTACCAGAGTTGGGGAACA 180

Qy 181 ttggtcgcaaggttcttctgattgttaaacactacagacagaatcacctcctctggatg 240
Db 181 TTGGTCTGAAGGTTCTTCTGATTGTAACTACAGACAGAAGATCACCTCTCTGGATGG 240

Qy 241 agccgatagtcaggttcccgggggccgtggagcgcgaacctatctcgtgtggtgg 300
Db 241 AGCCGATAGTCAAGTTCCTGGGGGCGGTGGAGCGCGCAACCTATATCTGTGTGATGGTGG 300

Qy 301 atccagatgcccctagcagacagaaacccagacagagattctggagacatitggctggtaa 360
Db 301 ATCCAGATGCCCTAGCAGACAGAAACCCACAGAGATTCTGGAGACATTTGGCTGGTAA 360

Qy 361 cagatatcaaggggocccagctgaagaaggaagattcaggggcagaggtatacagcct 420
Db 361 CAGATATCAAGGGGCCAGCTTGAAGAAGGAGATTTCAGGGCCAGGAGTATCAGCCT 420

Qy 421 accaggtccctcccccagggcacacagtggttccatcgctacagttcttctctatc 480
Db 421 ACCAGGCTCCCTCCCCAGGGCACACAGTGGCTTCCATCGCTACAGTTCCTTGTCTATC 480

Qy 481 ttccaggaaggaaggtctctctctcccaagggaacaaactcgaggtcttttggga 540
Db 481 TTCAGGAGGAAGGTCTCTCTCTCTCCCAAGGAACAACTCGAGGCTCTTGA 540

Qy 541 aaatggacagatttctgaacogttccacctggggggaacctgaagcaagcaccagttca 600
Db 541 AAATGGACAGATTCTGAACOGTTTCCACCTGGGGGAACCTGAAGCAAGCACCCAGTTCA 600

Qy 601 tgaccagaactaccagactaccacccctccagctccagagaaagggccagcgagc 660
Db 601 TGACCAGAACTACCAGACTACCACCCCTCCAGGCTCCAGAGAAAGGGCCAGCGAGC 660

Qy 661 ccaagcaaaaaaccaggcggagatagctgcctgtagtagcggcttggccatccggg 720
Db 661 CCAAGCAAAAAACCAGGCGGAGATAGTGCCTGCTAGATAGCCGGCTTGCCATCCGGG 720

Qy 721 catgtggccacactgccaccaccagcagatgtgggtatggaaacccctcttgatcacagaa 780
Db 721 CATGTGGCCACACTGCCACCACCAGCAGATGTGGGTATGGAAACCCCTCTGTGATACAGAA 780

Qy 781 cccctcttttccaaataaaaaaaatcatccaaaaa 826
Db 781 CCCCTCTTTTCCAAATAAAAAAATCATCCAAAAA 826

```

```

RESULT 4
US-60-147-499-7
; Sequence 7: Application US/60147499
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/60/147,499
; CURRENT FILING DATE: 1999-08-05
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 7
; LENGTH: 826

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 15..695
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 15..80
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 8.5
; OTHER INFORMATION: seq AALLLGLMMVVTG/DE
US-60-147-499-7

Query Match 100.0%; Score 826; DB 84; Length 826;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 826; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaccagaggtgccatgggttggacaatgagctggttcacagcagcactgttactgggtc 60
Db 1 aaccagaggtgccatgggttggacaatgagctggttcacagcagcactgttactgggtc 60

Qy 61 tcatgatggttgcactgagacagaggtgagaacagcccgctgtgccatgagccctcc 120
Db 61 tcatgatggttgcactgagacagaggtgagaacagcccgctgtgccatgagccctcc 120

Qy 121 tggcagagacacacctcttttggcagggccttgaagttttaccagagttggggaaca 180
Db 121 tggcagagacacacctcttttggcagggccttgaagttttaccagagttggggaaca 180

Qy 181 ttggtcgcaaggttcttctgattgttaaacactacagacagaatcacctcctctggatg 240
Db 181 ttggtcgcaaggttcttctgattgttaaacactacagacagaatcacctcctctggatg 240

Qy 241 agccgatagtcaggttcccgggggccgtggagcgcgaacctatctcgtgtggtgg 300
Db 241 agccgatagtcaggttcccgggggccgtggagcgcgaacctatctcgtgtggtgg 300

Qy 301 atccagatgcccctagcagacagaaacccagacagagattctggagacatitggctggtaa 360
Db 301 atccagatgcccctagcagacagaaacccagacagagattctggagacatitggctggtaa 360

Qy 361 cagatatcaaggggocccagctgaagaaggaagattcaggggcagaggtatacagcct 420
Db 361 cagatatcaaggggocccagctgaagaaggaagattcaggggcagaggtatacagcct 420

Qy 421 accaggtccctcccccagggcacacagtggttccatcgctacagttcttctctatc 480
Db 421 accaggtccctcccccagggcacacagtggttccatcgctacagttcttctctatc 480

Qy 481 ttccaggaaggaaggtctctctctcccaagggaacaaactcgaggtcttttggga 540
Db 481 ttccaggaaggaaggtctctctctcccaagggaacaaactcgaggtcttttggga 540

Qy 541 aaatggacagatttctgaacogttccacctggggggaacctgaagcaagcaccagttca 600
Db 541 aaatggacagatttctgaacogttccacctggggggaacctgaagcaagcaccagttca 600

Qy 601 tgaccagaactaccagactaccacccctccagctccagagaaagggccagcgagc 660
Db 601 tgaccagaactaccagactaccacccctccagctccagagaaagggccagcgagc 660

Qy 661 ccaagcaaaaaaccaggcggagatagctgcctgtagtagcggcttggccatccggg 720
Db 661 ccaagcaaaaaaccaggcggagatagctgcctgtagtagcggcttggccatccggg 720

Qy 721 catgtggccacactgccaccaccagcagatgtgggtatggaaacccctcttgatcacagaa 780
Db 721 catgtggccacactgccaccaccagcagatgtgggtatggaaacccctcttgatcacagaa 780

Qy 781 cccctcttttccaaataaaaaaaatcatccaaaaa 826
Db 781 cccctcttttccaaataaaaaaaatcatccaaaaa 826

```

RESULT 5  
US-60-169-629-7  
; Sequence 7, Application US/60169629  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Bougueleret, L.  
; APPLICANT: Jobert, S.  
; TITLE OF INVENTION: cDNAs for Secreted Proteins  
; FILE REFERENCE: GENSET.071PRF  
; CURRENT APPLICATION NUMBER: US/60/169,629  
; CURRENT FILING DATE: 1999-12-08  
; NUMBER OF SEQ ID NOS: 715  
; SOFTWARE: Patent.pm  
; SEQ ID NO 7  
; LENGTH: 826  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 15..695  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 15..80  
; OTHER INFORMATION: Von Heijne matrix  
; OTHER INFORMATION: score 8.5  
; OTHER INFORMATION: seq AALLGLMMVVTG/DE  
US-60-169-629-7

Query Match 100.0%; Score 826; DB 86; Length 826;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 826; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 aaccagaggtgccatgggttgacaatgaggtgttcacagcagcactgttactgggtc 60  
|||||  
DB 1 aaccagaggtgccatgggttgacaatgaggtgttcacagcagcactgttactgggtc 60  
QY 61 tcatgatgttggtcactgagacagagatgagaaaccccggtgtgccatgagggccctcc 120  
|||||  
DB 61 tcatgatgttggtcactgagacagagatgagaaaccccggtgtgccatgagggccctcc 120  
QY 121 tggacagagacacccctctttccaggccgttgagtttaccacagagttgggaaca 180  
|||||  
DB 121 tggacagagacacccctctttccaggccgttgagtttaccacagagttgggaaca 180  
QY 181 ttggctgcaaggtgttctctgattgttaacaaactacagacagagatcacctctggatgg 240  
|||||  
DB 181 ttggctgcaaggtgttctctgattgttaacaaactacagacagagatcacctctggatgg 240  
QY 241 agccgatgtcaagttcccgaggccgttgagcgcgaacctatattcctgggtgagtggtgg 300  
|||||  
DB 241 agccgatgtcaagttcccgaggccgttgagcgcgaacctatattcctgggtgagtggtgg 300  
QY 301 atccagatgccctacgacagacagacacccagacagagattctggagacattggctggtaa 360  
|||||  
DB 301 atccagatgccctacgacagacagacacccagacagagattctggagacattggctggtaa 360  
QY 361 cagatatcaagggtccgacccctgaagaaagggaagattcaggggcaggaggttatcagcct 420  
|||||  
DB 361 cagatatcaagggtccgacccctgaagaaagggaagattcaggggcaggaggttatcagcct 420  
QY 421 accaggtccctcccccacagcacacagtggttccatgcgtaccagttcttctgtctatc 480  
|||||  
DB 421 accaggtccctcccccacagcacacagtggttccatgcgtaccagttcttctgtctatc 480  
QY 481 ttcaggaaggaaggtcatctctctctctcccaaggaaacaaacacagagctcttgga 540  
|||||  
DB 481 ttcaggaaggaaggtcatctctctctctcccaaggaaacaaacacagagctcttgga 540  
QY 541 aaatggacagatttctgaacccgtttccacctgggggaacctgaagcaaacacccagttca 600  
|||||

Db 541 aaatggacagatttctgaacccgtttccacctggggcgaaacctgaagcaaacacccagttca 600  
QY 601 tgaccagaaactaccagagactaccacacccctccaggctcccccagagaaaagggccagcgagc 660  
|||||  
Db 601 tgaccagaaactaccagagactaccacacccctccaggctcccccagagaaaagggccagcgagc 660  
QY 661 ccaagcacaacaaacccaggcggagatagctgctgctagatagccggctttgcatcccggg 720  
|||||  
Db 661 ccaagcacaacaaacccaggcggagatagctgctgctagatagccggctttgcatcccggg 720  
QY 721 catgtggccacactgcccacacacagcagatgtgggtatggaaacccctctggatacagaa 780  
|||||  
Db 721 catgtggccacactgcccacacacagcagatgtgggtatggaaacccctctggatacagaa 780  
QY 781 cccctcttttccaaataaaaaaaatcatccaaaaaaataaaaaa 826  
|||||  
Db 781 cccctcttttccaaataaaaaaaatcatccaaaaaaataaaaaa 826  
RESULT 6  
US-60-187-470-7  
; Sequence 7, Application US/60187470  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Bougueleret, L.  
; APPLICANT: Jobert, S.  
; TITLE OF INVENTION: cDNAs for Secreted Proteins  
; FILE REFERENCE: 78.US2.PRO  
; CURRENT APPLICATION NUMBER: US/60/187,470  
; CURRENT FILING DATE: 2000-03-06  
; NUMBER OF SEQ ID NOS: 715  
; SOFTWARE: Patent.pm  
; SEQ ID NO 7  
; LENGTH: 826  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 15..695  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 15..80  
; OTHER INFORMATION: Von Heijne matrix  
; OTHER INFORMATION: score 8.5  
; OTHER INFORMATION: seq AALLGLMMVVTG/DE  
US-60-187-470-7

Query Match 100.0%; Score 826; DB 93; Length 826;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 826; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 aaccagaggtgccatgggttgacaatgaggtgttcacagcagcactgttactgggtc 60  
|||||  
Db 1 aaccagaggtgccatgggttgacaatgaggtgttcacagcagcactgttactgggtc 60  
QY 61 tcatgatgttggtcactgagacagagatgagaaaccccggtgtgccatgagggccctcc 120  
|||||  
Db 61 tcatgatgttggtcactgagacagagatgagaaaccccggtgtgccatgagggccctcc 120  
QY 121 tggacagagacacccctctttccaggccgttgagtttaccacagagttgggaaca 180  
|||||  
Db 121 tggacagagacacccctctttccaggccgttgagtttaccacagagttgggaaca 180  
QY 181 ttggctgcaaggtgttctctgattgttaacaaactacagacagagatcacctctggatgg 240  
|||||  
Db 181 ttggctgcaaggtgttctctgattgttaacaaactacagacagagatcacctctggatgg 240  
QY 241 agccgatgtcaagttcccgaggccgttgagcgcgaacctatattcctgggtgagtggtgg 300  
|||||  
Db 241 agccgatgtcaagttcccgaggccgttgagcgcgaacctatattcctgggtgagtggtgg 300  
QY 301 atccagatgccctacgacagacagacacccagacagagattctggagacattggctggtaa 360  
|||||



Db 301 atccagatgcctccctagcagagcagaccacacagacagagattctggagacattggctggttaa 360  
Qy 361 cagatatcaaggccgcacccacctgaaagaaggaaagattcagggccagagttatcagcct 420  
Db 361 cagatatcaaggccgcacccacctgaaagaaggaaagattcagggccagagttatcagcct 420  
Qy 421 accaggtccctcccccacccgacacagtggtttccatcgctaccagttctttgtctatc 480  
Db 421 accaggtccctcccccacccgacacagtggtttccatcgctaccagttctttgtctatc 480  
Qy 481 ttccaggaaggaaagtgatctctctctcccccaggaagaaacacacactcgaggtctttgga 540  
Db 481 ttccaggaaggaaagtgatctctctctcccccaggaagaaacacacactcgaggtctttgga 540  
Qy 541 aaatgagcagatttctgaacgtttccacgttccacgttggcggaacctgaaagcagcaccagttca 600  
Db 541 aaatgagcagatttctgaacgtttccacgttccacgttggcggaacctgaaagcagcaccagttca 600  
Qy 601 tgaccagaactaccagagactcaccacccctccaggtctccagagaaaggccagcgagc 660  
Db 601 tgaccagaactaccagagactcaccacccctccaggtctccagagaaaggccagcgagc 660  
Qy 661 ccaagcacaacacccagcgagagatgctgctgctagatagccggtttgccatccggg 720  
Db 661 ccaagcacaacacccagcgagagatgctgctgctagatagccggtttgccatccggg 720  
Qy 721 catgtggccacactgccacacccagcagatgtgggtatggaacccctctgtgatacagaa 780  
Db 721 catgtggccacactgccacacccagcagatgtgggtatggaacccctctgtgatacagaa 780  
Qy 781 cccctttttcccaataaaaaaaatccatcccaaaaaaa 826  
Db 781 cccctttttcccaataaaaaaaatccatcccaaaaaaa 826

RESULT 7  
US-60-172-373-10894  
; Sequence 10894, Application US/60172373  
; GENERAL INFORMATION:  
; APPLICANT: Morris, MacDonald  
; APPLICANT: Lal, Preeti  
; APPLICANT: Diep, Dinh  
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using  
; FILE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polymorph  
; FILE REFERENCE: GX-0006 P  
; CURRENT APPLICATION NUMBER: US/60/172,373  
; CURRENT FILING DATE: 1999-12-16  
; NUMBER OF SEQ ID NOS: 25,772  
; SOFTWARE: PERL Program  
; SEQ ID NO 10894  
; LENGTH: 904  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No: 204392.1  
US-60-172-373-10894

Query Match 73.8%; Score 610; DB 87; Length 904;  
Best Local Similarity 99.5%; Pred. No. 4.9e-276;  
Matches 810; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 aaccagagtgccatgggttgacacagtggtgacacagcagcactgttactgggtc 60  
Db 75 aaccagagtgccatgggttgacacagtggtgacacagcagcactgttactgggtc 134  
Qy 61 tcatgatgtgttcactgagacagagatgagaacagccgtgtgcccattgagggccctcc 120  
Db 135 tcatgatgtgttcactgagacagagatgagaacagccgtgtgcccattgagggccctct 194  
Qy 121 tggacgaggaacacccctcttttgcagggccttgaaagttttctaccacagagttggggaaca 180

Db 195 tggacgaggaacacccctcttttgcagggccttgaaagttttctaccacagagttggggaaca 254  
Qy 181 ttggctgcaaggttgttctctgattgtaacaactacagacagaaatcacctcctctggtg 240  
Db 255 ttggctgcaaggttgttctctgattgtaacaactacagacagaaatcacctcctctggtg 314  
Qy 241 agccgatagtcaagtctccggggcggtggacggcgcaacctatatctctggtggtg 300  
Db 315 agccgatagtcaagtctccggggcggtggacggcgcaacctatatctctggtggtg 374  
Qy 301 atccagatgcctccctagcagagacagaaacccacagagattctggagacattggcctggttaa 360  
Db 375 atccagatgcctccctagcagagacagaaacccacagagattctggagacattggcctggttaa 434  
Qy 361 cagatatcaaggccgcacccacctgaaagaaggaaagattcagggccagagttatcagcct 420  
Db 435 cagatatcaaggccgcacccacctgaaagaaggaaagattcagggccagagttatcagcct 494  
Qy 421 accaggtccctcccccacccgacacagtggtttccatcgctaccagttctttgtctatc 480  
Db 495 accaggtccctcccccacccgacacagtggtttccatcgctaccagttctttgtctatc 554  
Qy 481 ttccaggaaggaaagtgatctctctctcccccaggaagaaacacacactcgaggtctttgga 540  
Db 555 ttccaggaaggaaagtgatctctctctcccccaggaagaaacacacactcgaggtctttgga 614  
Qy 541 aaatgagcagatttctgaacgtttccacgttccacgttggcggaacctgaaagcagcaccagttca 600  
Db 615 aaatgagcagatttctgaacgttccacgttggcggaacctgaaagcagcaccagttca 674  
Qy 601 tgaccagaactaccagagactcaccacccctccaggtctccagagaaaggccagcgagc 660  
Db 675 tgaccagaactaccagagactcaccacccctccaggtctccagagaaaggccagcgagc 734  
Qy 661 ccaagcacaacacccagcgagagatgctgctgctagatagccggtttgccatccggg 720  
Db 735 ccaagcacaacacccagcgagagatgctgctgctagatagccggtttgccatccggg 794  
Qy 721 catgtggccacactgccacacccagcagatgtgggtatggaacccctctgtgatacagaa 780  
Db 795 catgtggccacactgccacacccagcagatgtgggtatggaacccctctgtgatacagaa 854  
Qy 781 cccctttttcccaataaaaaaaatccatcccaaaaaaa 814  
Db 855 cccctttttcccaataaaaaaaatccatcccaaaaaaa 888

RESULT 8  
PCT-US99-29950-56  
; Sequence 56, Application PC/TUS9929950  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: 47 Human Secreted Proteins  
; FILE REFERENCE: P2035.PCT  
; CURRENT APPLICATION NUMBER: PCT/US99/29950  
; CURRENT FILING DATE: 1999-12-16  
; EARLIER APPLICATION NUMBER: 60/112,809  
; EARLIER FILING DATE: 1998-12-17  
; EARLIER APPLICATION NUMBER: 60/113,006  
; EARLIER FILING DATE: 1998-12-18  
; NUMBER OF SEQ ID NOS: 230  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 56  
; LENGTH: 1028  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (1022)  
; OTHER INFORMATION: n equals a,t,g, or c  
; FEATURE:  
; NAME/KEY: SITE

; LOCATION: (1026)  
; OTHER INFORMATION: n equals a,t,g, or c  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (1027)  
; OTHER INFORMATION: n equals a,t,g, or c  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (1028)  
; OTHER INFORMATION: n equals a,t,g, or c  
PCT-US99-29950-56

Query Match 73.8%; Score 610; DB 1; Length 1028;  
Best Local Similarity 99.5%; Pred. No. 4.9e-276;  
Matches 810; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 aaccagagtgcccatgggttgacaatgagctgtgtcacagcagcactgttactgggtc 60  
Db 139 aaccagagtgcccatgggttgacaatgagctgtgtcacagcagcactgttactgggtc 198  
QY 61 tcattgatggtgctcactggagacagagatgagacagcccgctgtgccatgagccctcc 120  
Db 199 tcattgatggtgctcactggagacagagatgagacagcccgctgtgccatgagccctct 258  
QY 121 tggacagagacacccctcttttgcaggcccttgaagtcttaccacagagtggtgggaaca 180  
Db 259 tggacagagacacccctcttttgcaggcccttgaagtcttaccacagagtggtgggaaca 318  
QY 181 ttggctgcagggtgttctctgattgaacacacacagacagagatcactctctgtagg 240  
Db 319 ttggctgcagggtgttctctgattgaacacacacagacagagatcactctctgtagg 378  
QY 241 agccgatagtcgaagtcccggtggccggtggacggcgcaacctatctcctgggtgtaggtg 300  
Db 379 agccgatagtcgaagtcccggtggccggtggacggcgcaacctatctcctgggtgtaggtg 438  
QY 301 atccagatgcccctagcagacagaacccacagacagagattcttgagacattggtggttaa 360  
Db 439 atccagatgcccctagcagacagaacccacagacagagattcttgagacattggtggttaa 498  
QY 361 cagatatcaagggtgcgcagcctgaagaaagggaagattcaggcccgaggagttatcagcct 420  
Db 499 cagatatcaagggtgcgcagcctgaagaaagggaagattcaggcccgaggagttatcagcct 558  
QY 421 accaggtccctccacacggcacagtggttccatgctaccaggttcttctgtctatc 480  
Db 559 accaggtccctccacacggcacagtggttccatgctaccaggttcttctgtctatc 618  
QY 481 ttcaggaaggaaaggtcatctctctctcccaaggaaacaaacacacacacacacacacacac 540  
Db 619 ttcaggaaggaaaggtcatctctctctcccaaggaaacaaacacacacacacacacacacac 678  
QY 541 aaatgacagagattctgaacggtttccacgtggcgcaacctgaagcaagcaccagttca 600  
Db 679 aaatgacagagattctgaacggtttccacgtggcgcaacctgaagcaagcaccagttca 738  
QY 601 tgaccacgaactccaggactccacacccctccaggtccacagagaaaggccagcgagc 660  
Db 739 tgaccacgaactccaggactccacacccctccaggtccacagagaaaggccagcgagc 798  
QY 661 ccaagcacaaaaaccaggcgagatagctgctgctgctagatagcggctttgcatccggg 720  
Db 799 ccaagcacaaaaaccaggcgagatagctgctgctgctgctagatagcggctttgcatccggg 858  
QY 721 catgtggcacactgcccaccccgagctgtgggtatggaacccctctggatcacagaa 780  
Db 859 catgtggcacactgcyaccaccccgagctgtgggtatggaacccctctggatcacagaa 918  
QY 781 cccctctttttccaaataaaaaaaatcatcca 814  
Db 919 cccctctttttccaaataaaaaaaatcatcca 952

RESULT 9

US-09-359-922-5157  
; Sequence 5157, Application US/09359922  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA  
; FILE REFERENCE: 20411-752CON1  
; CURRENT APPLICATION NUMBER: US/09/359,922  
; CURRENT FILING DATE: 1999-07-22  
; EARLIER APPLICATION NUMBER: US 09/205,155  
; EARLIER FILING DATE: 1998-12-03  
; NUMBER OF SEQ ID NOS: 13203  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5157  
; LENGTH: 878  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-359-922-5157

Query Match 67.7%; Score 559; DB 50; Length 878;  
Best Local Similarity 99.4%; Pred. No. 4.4e-252;  
Matches 809; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 aaccagagtgcccatgggttgacaatgagctgtgtcacagcagcactgttactgggtc 60  
Db 57 aaccagagtgcccatgggttgacaatgagctgtgtcacagcagcactgttactgggtc 116  
QY 61 tcattgatggtgctcactggagacagagatgagacagcccgctgtgccatgagccctcc 120  
Db 117 tcattgatggtgctcactggagacagagatgagacagcccgctgtgccatgagccctct 176  
QY 121 tggacagagacacccctcttttgcaggcccttgaagtcttaccacagagtggtgggaaca 180  
Db 177 tggacagagacacccctcttttgcaggcccttgaagtcttaccacagagtggtgggaaca 236  
QY 181 ttggctgcagggtgttctctgattgaacacacacagacagagatcactctctgtaggtg 240  
Db 237 ttggctgcagggtgttctctgattgaacacacacagacagagatcactctctgtaggtg 296  
QY 241 agccgatagtcgaagtcccggtggccggtggacggcgcaacctatctcctgggtgtaggtg 300  
Db 297 agccgatagtcgaagtcccggtggccggtggacggcgcaacctatctcctgggtgtaggtg 356  
QY 301 atccagatgcccctagcagacagaacccacagacagattcttgagacattggtggttaa 360  
Db 357 atccagatgcccctagcagacagaacccacagacagattcttgagacattggtggttaa 416  
QY 361 cagatatcaagggtgcgcagcctgaagaaagggaagattcaggcccgaggagttatcagcct 420  
Db 417 cagatatcaagggtgcgcagcctgaagaaagggaagattcaggcccgaggagttatcagcct 476  
QY 421 accaggtccctccacacggcacagtggttccatgctaccaggttcttctgtctatc 480  
Db 477 accaggtccctccacacggcacagtggttccatgctaccaggttcttctgtctatc 536  
QY 481 ttcaggaaggaaaggtcatctctctctcccaaggaaacaaacacacacacacacacacacac 540  
Db 537 ttcaggaaggaaaggtcatctctctctcccaaggaaacaaacacacacacacacacacacac 596  
QY 541 aaatgacagagattctgaacccgtttccacgtggcgcaacctgaagcaagcaccagttca 600  
Db 597 aaatgacagagattctgaacccgtttccacgtggcgcaacctgaagcaagcaccagttca 656  
QY 601 tgaccacgaactccaggactccacacccctccaggtccacagagaaaggccagcgagc 660  
Db 657 tgaccacgaactccaggactccacacccctccaggtccacagagaaaggccagcgagc 716  
QY 661 ccaagcacaaaaaccaggcgagatagctgctgctgctagatagcggctttgcatccggg 720  
Db 717 ccaagcacaaaaaccaggcgagatagctgctgctgctgctagatagcggctttgcatccggg 776

QY 721 catgtggcacactgcccaccaccgacgatgtgggtatggaaacccctctgtgatacagaa 780  
Db 777 catgtggcacactgcccaccaccgacgatgtgggtatggaaacccctctgtgatacagaa 836  
QY 781 cccctcttttccaaataaaaaaaatcatcca 814  
Db 837 cccctcttttccaaataaaaaaaatcatcca 870

RESULT 10  
US-09-359-922-5157  
; Sequence 5157, Application US/09359922A  
; GENERAL INFORMATION:  
; APPLICANT: Leshkowitz, Dena  
; APPLICANT: Liu, Jin  
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA  
; FILE REFERENCE: 20411-752CON1  
; CURRENT APPLICATION NUMBER: US/09/359,922A  
; CURRENT FILING DATE: 1999-07-22  
; EARLIER APPLICATION NUMBER: US 09/205,155  
; EARLIER FILING DATE: 1998-12-03  
; EARLIER APPLICATION NUMBER: US 09/034,341  
; EARLIER FILING DATE: 1998-02-13  
; NUMBER OF SEQ ID NOS: 13203  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5157  
; LENGTH: 878  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-359-922-5157

Query Match 67.7%; Score 559; DB 92; Length 878;  
Best Local Similarity 99.4%; Pred. No. 4.4e-252;  
Matches 809; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 aaccagaggtgccatgggttgacaatgaggtggtcacagcagcactgttactgggtc 60  
Db 57 aaccagaggtgccatgggttgacaatgaggtggtcacagcagcactgttactgggtc 116  
QY 61 tcattgattggtcactggagacagagatgagaaacccgtgtgcccattgagggccctcc 120  
Db 117 tcattgattggtcactggagacagagatgagaaacccgtgtgcccattgagggccctct 176  
QY 121 tgacagagacacccctcttttgcaggcccttgaaagtcttctaccagagattggggaaca 180  
Db 177 tgacagagacacccctcttttgcaggcccttgaaagtcttctaccagagattggggaaca 236  
QY 181 ttggtcgaaggttggtctctgattgttaacaactacagacagagatcacctcctggtg 240  
Db 237 ttggtcgaaggttggtctctgattgttaacaactacagacagagatcacctcctggtg 296  
QY 241 accgatagtcagttctccggggccgtgacggcccaacctatctcctggtgattggtg 300  
Db 297 agccgatgctcagttctccggggccgtgacggcccaacctatctcctggtgattggtg 356  
QY 301 atccagatgcccctagcagacgacacccagacagagattctggagacattggtggttaa 360  
Db 357 atccagatgcccctagcagacgacacccagacagagattctggagacattggtggttaa 416  
QY 361 cagatatcaaggcgccgacctgaagaaagggaagattccaggccaggagattatcagcct 420  
Db 417 cagatatcaaggcgccgacctgaagaaagggaagattccaggccaggagattatcagcct 476  
QY 421 accaggtcctcccccaccggacacagatggttccatcagctaccagttcttctgtctatc 480  
Db 477 accaggtcctcccccaccggacacagatggttccatcagctaccagttcttctgtctatc 536  
QY 481 ttcaggaaggaagggtctatctctctctctcccaaggaaacaaacacagggctcttggga 540  
Db 537 ttcaggaaggaagggtctatctctctctctcccaaggaaacaaacacagggctcttggga 596

QY 541 aaatggacagattcttgaaaccttttccacctgtggggaacctgaagcaagcaccagttca 600  
Db 597 aaatggacagattcttgaaaccttttccacctgtggggaacctgaagcaagcaccagttca 656  
QY 601 tgaccacgaactaccaggaactcaccacacctccaggtccacagagaaggccagcgagc 660  
Db 657 tgaccacgaactaccaggaactcaccacacctccaggtccacagagaaggccagcgagc 716  
QY 661 ccaagcacaaaaaccaggcgggagatgctgctgtagatagccgcttttgcctatccggg 720  
Db 717 ccaagcacaaaaaccaggcgggagatgctgctgtagatagccgcttttgcctatccggg 776  
QY 721 catgtggccacactgcccaccaccgacgatgtgggtatggaaacccctctgtgatacagaa 780  
Db 777 catgtggccacactgcccaccaccgacgatgtgggtatggaaacccctctgtgatacagaa 836  
QY 781 cccctcttttccaaataaaaaaaatcatcca 814  
Db 837 cccctcttttccaaataaaaaaaatcatcca 870

RESULT 11  
US-08-958-820-4  
; Sequence 4, Application US/08958820  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer  
; APPLICANT: Corley, Neil  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: HUMAN PHOSPHOLIPID BINDING PROTEINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/958,820  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0379 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 903 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: LUNGUT12  
; CLONE: 3126479  
US-08-958-820-4

Query Match 67.7%; Score 559; DB 29; Length 903;  
Best Local Similarity 99.4%; Pred. No. 4.4e-252;  
Matches 809; Conservative 0; Mismatches 5; Indels 0; Gaps 0;



QY 541 aaatggacagatttctgaacccgtttccacccctggcggaacctgaagcagaccaccagttca 600  
 DB 614 AAATGGACAGATTCTTGAACCGCTTCCACCTGGCGGAACCTTGAAGCAAGCAACCCAGTTCA 673  
 QY 601 tgaccagaactacagagactcaccacccctccaggtcccccagagaaaggccagcagc 660  
 DB 674 TGACCCAGAACTACAGAGACTACCAACCCCTCCAGGCTCCAGAGAGAGAGGCCAGCAGC 733  
 QY 661 ccaagcaaaaaaacaggcgagatagctgctgctagatagccggcgtttgccatccggg 720  
 DB 734 CCAAGCACAACCAAGCGGAGATAGCTGCTGCTAGATAGCGGCTTTGCCATCCGGG 793  
 QY 721 catgtggccacactgcccaccacagcagatgtgggtatggaaacccctctctgtgatacagaa 780  
 DB 794 CATGTGGCCACACTGCCACCAAGCAGCATGTGGGTATGGAAACCCCTCTCTGGATACAGAA 853  
 QY 781 cccctctcttccaaataaaaaaaatcatcca 814  
 DB 854 CCCCTCTCTTCCAAATAAAAAATCATCCA 887

RESULT 13  
 US-09-390-126-4  
 ; Sequence 4, Application US/09390126  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lal, Preeti  
 ; APPLICANT: Hillman, Jennifer  
 ; APPLICANT: Corley, Neil  
 ; APPLICANT: Shah, Purvi  
 ; TITLE OF INVENTION: HUMAN PHOSPHOLIPID BINDING PROTEINS  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Dr.  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/390,126  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/958,820  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Billings, Lucy J.  
 ; REGISTRATION NUMBER: 36,749  
 ; REFERENCE/DOCKET NUMBER: PF-0379 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 650-855-0555  
 ; TELEFAX: 650-845-4166  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 903 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: LUNGTUT12  
 ; CLONE: 3126479  
 ; US-09-390-126-4

Query Match 67.7%; Score 559; DB 92; Length 903;  
 Best Local Similarity 99.4%; Pred. No. 4.4e-252;  
 Matches 809; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 aaccagaggtgccccatgggttggaacatgaggtggttcacagcagcagcactgttactgggtc 60  
 DB 74 AACACAGAGGTGCCCATGGGTGGACAAATGAGGCTGGTGCACAGCAGCACTGTTACTGGGTC 133  
 QY 61 tcatgatgggtggttcaactgagagagagatgagaacacgcccgtgtgcccagagccctcc 120  
 DB 134 TCATGATGGTGGTCACTGGAGAGAGAGATGAGAACACCCCGTGTGCCCATGAGGCCCTCT 193  
 QY 121 tggagcagagacacccctcttttggcagggcccttgaagtttttaccacagagtggggaaaca 180  
 DB 194 TGGACGAGGACACCTCTTTTCCAGGGCCTTGAAGTTTCTTACCAGAGTTGGGAACA 253  
 QY 181 ttgctcgaaggtgttctcctgattgttaacactacacagagaagatcacctcctcctggatgg 240  
 DB 254 TTGCTCTCAAGGTGTCTCTGATTGTAACTACACAGACAGATTCCTGGATGG 313  
 QY 241 agccgatagtcgaagttcccgggccgtgacgcgcgaacctatctctgtgtggtggtg 300  
 DB 314 AGCCGATAGTCAAGTTCCTGGGGCCCGTGGACGGCGCAACCTATATCTGGTGTGTTGG 373  
 QY 301 atccagatgccccctagcagagcagacccacagagagattcttgagacattgggtggttaa 360  
 DB 374 ATCCAGATGCCCTTAGCAGAGCAGACCCACAGAGATTCCTGGAGACATTGGCTGGTAA 433  
 QY 361 cagatatcaagggcgccgacctgaagaaggaagattcaggccagaggtatcagcct 420  
 DB 434 CAGATATCAAGGGCGCCGACCTGAAGGAAGGAAGATTTCAGGGCCAGAGTTATCAGCCT 493  
 QY 421 accaggtccctcccccacggcacacagtggtctccatcgctaccagttcttctctatc 480  
 DB 494 ACCAGGTCTCTCCACCGGCACACAGTGGCTTCCATCGCTACCATCTCTTGTCTATC 553  
 QY 481 ttcaggaaggaaggtctctctctctcccaaggaaaaacaaactcagaggtctcttggga 540  
 DB 554 TTCAGGAAGGAAAAAGTCTCTCTCTCCAAAGGAAAAACAAACTCGAGGCTCTTGA 613  
 QY 541 aaatggacagatttctgaacccgtttccacccctggcggaacctgaagcagaccaccagttca 600  
 DB 614 AAATGGACAGATTCTTGAACCGCTTCCACCTGGCGGAACCTTGAAGCAAGCAACCCAGTTCA 673  
 QY 601 tgaccagaactacagagactcaccacccctccaggtcccccagagaaaggccagcagc 660  
 DB 674 TGACCCAGAACTACAGAGACTCACCACCTTCCAGGCTCCAGAGAGAGGCCAGCAGC 733  
 QY 661 ccaagcaaaaaaacaggcgagatagctgctgctagatagccggcgtttgccatccggg 720  
 DB 734 CCAAGCACAACCAAGCGGAGATAGCTGCTGCTAGATAGCGGCTTTGCCATCCGGG 793  
 QY 721 catgtggccacactgcccaccacagcagatgtgggtatggaaacccctctgatacagaa 780  
 DB 794 CATGTGGCCACACTGCCACCAAGCAGCATGTGGGTATGGAAACCCCTCTCTGGATACAGAA 853

RESULT 14  
 US-09-390-126-4  
 ; Sequence 4, Application US/09390126  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lal, Preeti  
 ; APPLICANT: Hillman, Jennifer  
 ; APPLICANT: Corley, Neil  
 ; APPLICANT: Shah, Purvi  
 ; TITLE OF INVENTION: HUMAN PHOSPHOLIPID BINDING PROTEINS  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Dr.  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA

ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/390,126  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/958,820  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0379 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 903 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: LUNGTUT12  
CLONE: 3126479  
US-09-390-126-4

Query Match 57.7%; Score 559; DB 92; Length 903;  
Best Local Similarity 99.4%; Pred. No. 4.4e-252;  
Matches 809; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 aaccagaggtgcccattgggtgacaaatgagctggtcagacagcactgtactgggtc 60  
DB 74 AACGAGAGTGGCCATGGGTGGACATGAGGCTGTCTACGAGCAGCCTTACTGGGTC 133  
QY 61 tcatgattggtgctactgagacgaggtgagaaacgcccgtgtgcccattgagccctcc 120  
DB 134 TCATGATGTTGTTGCTGATGTTGAACACTACAGACAGAGATCACCCTCTGGATGG 193  
QY 121 tggacagagacacccctctttccagggccttgaactttctaccagagttggggaaca 180  
DB 194 TGGACGAGACACCCCTCTTTGCCAGGGCCCTTGAAGTTTCTACCCAGAGTTGGGGAACA 253  
QY 181 tbggctgcaaggtgttctctgattgttaacaactacagacagaatcacctctctgattgg 240  
DB 254 TTGGCTGCAAGGTTGTTCTGATGTTGAACACTACAGACAGAGATCACCCTCTGGATGG 313  
QY 241 agccgagatcaagttcccgggggccgtggagcggcgcaacctatatctctggtgattgg 300  
DB 314 AGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAACCTATATCTCTGATGGTGG 373  
QY 301 atccagatgcccctagcagacagacaccagacagagattctgagacattgctggtgtaa 360  
DB 374 ATCCAGATGCCCTAGCAGACGACAGACACCAGACAGAGATCTGGAGACATGTCGTGTA 433  
QY 361 cagatacaagggcgccgacctgaagaaaggaagattcaggggccaggagttatcagcct 420  
DB 434 CAGATATCAAGGGCGCCGACCTGAAGGAAGGAAGATTTCAGGGCCAGGAGTTATCAGCT 493  
QY 421 accaggtccctcccccacgacacagtggtcttcacgtgctaccagttcttctgtctatc 480  
DB 494 ACCAGGCTCCCTCCCGCCGACAGTGGCTTCCATGCGTACCAGTTCTTTGTCTATC 553  
QY 481 ttccaggaagaaaggtcatctctctctcccaaggaagaaacaaactcaggtcttggga 540  
DB 554 TTCAGGAGGAAAGTATCTCTCTCTCCATCCCAAGGAAACAACTCAGGCTCTTGA 613  
QY 541 aaatggacagattctgaaacggtttccacctgggcgaaacctgaagcaagcaccagttca 600

DB 614 AAATGACACAGATTTCTGAAACCGCTTCCACCTGGCGAACCTGAAGCAAGCACCAGTTCA 673  
QY 601 tgaccagaactaccagagactcacaacccctccaggtcccccagagaagggccagcgagc 660  
DB 674 TGACCCAGAACTACCAGGACTACCAACCCCTCCAGGCTCCAGAGAGGGCGCAGGAGC 733  
QY 661 ccaagcacaaaaaccagcgagagtagctcctgctagatagccggtttgcatcccg 720  
DB 734 CCACGACAAAAACAGCGGAGATAGCTGCTGTAGATAGCCGGCTTGGCATCCCGG 793  
QY 721 catgtggccacactgcccaccaccgacgatgtgggtatggaacccctcttgatacagaa 780  
DB 794 CATGTGGCCACACTGCCACACACCGACGATGFGGTATGGAACCCCTCTGGATACAGAA 853  
QY 781 ccccttttttccaaataaaaaaaatcatcca 814  
DB 854 CCCCTTCTTTTCCAAATAAAAAAATCATCCA 887

## RESULT 15

PCT-US99-29950-61  
; Sequence 61, Application PC/TUS9929950  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: 47 Human Secreted Proteins  
; FILE REFERENCE: P2035.PCT  
; CURRENT APPLICATION NUMBER: PCT/US99/29950  
; CURRENT FILING DATE: 1999-12-16  
; EARLIER APPLICATION NUMBER: 60/112,809  
; EARLIER FILING DATE: 1998-12-17  
; EARLIER APPLICATION NUMBER: 60/113,006  
; EARLIER FILING DATE: 1998-12-18  
; NUMBER OF SEQ ID NOS: 230  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 61  
; LENGTH: 952  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
PCT-US99-29950-61

Query Match 58.0%; Score 479; DB 1; Length 952;  
Best Local Similarity 99.8%; Pred. No. 1.7e-214;  
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 285 atccctggtgattggtggtatccagatgcccctagcagagcagacccagacagagattctgg 344  
DB 335 atccctggtgattggtggtatccagatgcccctagcagagcagacccagacagagattctgg 394  
QY 345 agacattgctggttaacagatatcaaggcgccgacctgaagaaaggaagattcagggc 404  
DB 395 agacattgctggttaacagatatcaaggcgccgacctgaagaaaggaagattcagggc 454  
QY 405 caggattatcagctaccagctccctcccccacacagtcagtcagtcagtcagtcagtcagtc 464  
DB 455 caggattatcagctaccagctccctcccccacacagtcagtcagtcagtcagtcagtcagtc 514  
QY 465 cagttcttctctatcttcaggaaaggaaggtcatctctctctcccaaggaagaaacaaa 524  
DB 515 cagttcttctctatcttcaggaaaggaaggtcatctctctctcccaaggaagaaacaaa 574  
QY 525 actcagagctcttggaataatggacagatttttgaaacgtttccacctgggccaacctgaa 584  
DB 575 actcagagctcttggaataatggacagatttttgaaacgtttccacctgggccaacctgaa 634  
QY 585 gcaagcaccagttcatgaccagacactaccagactcaccaacctccaggtccccaga 644  
DB 635 gcaagcaccagttcatgaccagacactaccagactcaccaacctccaggtccccaga 694  
QY 645 gaaagggccagcgagcccaagcacaacaaacccagcgagatagctgctgctagatagcc 704  
DB 695 gaaagggccagcgagcccaagcacaacaaacccagcgagatagctgctgctagatagcc 754

Qy	705	ggctttgcacatcgggcgatgtgccacactgcccacacacgcacgatgtgggtatggaacc	764
Db	755	ggctttgcacatcgggcgatgtgccacactgcccacacacgcacgatgtgggtatggaacc	814
Qy	765	ccctctggatcacgaacccctttttccaaataaaaaaaatcatcca	814
Db	815	ccctctggtacagaaacccctttttccaaataaaaaaaatcatcca	864

Search completed: May 1, 2000, 20:00:31  
Job time: 22015 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2000, 14:44:56 ; Search time 3022.95 Seconds  
(without alignments)  
1031.674 Million cell updates/sec

Title: US-09-215-435-124  
Perfect score: 826  
Sequence: 1 aaccagaggtgccatgggt.....atcatcccaaaaaaaaaa 826

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4538634 seqs, 1887831982 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1:	em_est1:*
2:	em_est2:*
3:	em_est3:*
4:	em_est4:*
5:	em_est5:*
6:	em_est6:*
7:	em_est7:*
8:	em_est8:*
9:	em_est9:*
10:	em_est10:*
11:	em_est11:*
12:	em_est12:*
13:	em_est13:*
14:	em_est14:*
15:	em_est15:*
16:	em_est16:*
17:	em_est17:*
18:	em_est18:*
19:	em_est19:*
20:	gb_est1:*
21:	gb_est2:*
22:	gb_est3:*
23:	gb_est4:*
24:	gb_est5:*
25:	gb_est6:*
26:	gb_est7:*
27:	gb_est8:*
28:	gb_est9:*
29:	gb_est10:*
30:	gb_est11:*
31:	gb_est12:*
32:	gb_est13:*
33:	gb_est14:*
34:	gb_est15:*
35:	gb_est16:*
36:	gb_est17:*
37:	gb_est18:*
38:	gb_est19:*
39:	gb_est20:*
40:	gb_est21:*
41:	gb_est22:*
42:	gb_est23:*
43:	gb_est24:*

44:	gb_est25:*
45:	gb_est26:*
46:	gb_est27:*
47:	gb_est28:*
48:	gb_est29:*
49:	gb_est30:*
50:	gb_est31:*
51:	gb_est32:*
52:	em_est20:*
53:	em_est21:*
54:	em_est22:*
55:	em_est23:*
56:	em_est24:*
57:	em_est25:*
58:	em_est26:*
59:	gb_est33:*
60:	gb_est34:*
61:	gb_est35:*
62:	gb_est36:*
63:	gb_est37:*
64:	gb_est38:*
65:	em_est27:*
66:	em_est28:*
67:	em_est29:*
68:	em_est30:*
69:	gb_est39:*
70:	gb_est40:*
71:	gb_est41:*
72:	gb_est42:*
73:	gb_est43:*
74:	gb_est44:*
75:	em_est31:*
76:	em_est32:*
77:	em_est33:*
78:	em_est34:*
79:	gb_gss1:*
80:	gb_gss2:*
81:	gb_gss3:*
82:	gb_gss4:*
83:	em_gss1:*
84:	em_gss2:*
85:	em_gss3:*
86:	em_gss4:*
87:	gb_gss5:*
88:	gb_gss6:*
89:	gb_gss7:*
90:	gb_gss8:*
91:	gb_gss9:*
92:	em_gss5:*
93:	em_gss6:*
94:	em_gss7:*
95:	em_gss8:*
96:	em_gss9:*
97:	em_gss10:*
98:	em_gss11:*
99:	gb_gss10:*
100:	gb_gss11:*
101:	em_gss12:*
102:	gb_gss12:*
103:	gb_gss13:*
104:	gb_gss14:*
105:	gb_gss15:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	376	45.5	552	43	AI218954
					AI218954 qg72q08.x

Seq primer: -40UP from Gibco	
High quality sequence stop: 423.	
Location/Qualifiers	
1. 552	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:1840766"	
/lab_host="DH10B"	
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with	
a modified polylinker; Site_1: Not I; Site_2: Eco RI;	
Equal amounts of plasmid DNA from three normalized	
libraries (fetal lung NDHL19W, testis NHT, and B-cell	
NCI_CGAP_GCB1) were mixed, and ss circles were made as	
tracer in a subtractive hybridization reaction. The driver	
was PCR-amplified cDNAs from pools of 5,000 clones made	
from the same 3 libraries. The pools consisted of	
I.M.A.G.E. clones 297480-302087, 682632-687239,	
726408-728711, and 729096-731399. Subtraction by Bento	
Soares and M. Fatima Bonaldo."	
BASE COUNT	98 a 129 c 158 g 167 t
ORIGIN	
Query Match 45.5%; Score 376; DB 43; Length 552;	
Best Local Similarity 99.8%; Pred. No. 4.9e-167;	
Matches 426; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Qy	388 aagggaagattcaggccaggaggtatcagctaccaggtccctcccccacggcacaca 447
Db	446 AAGGAAGATTTCAGGGCCAGGAGTTATCAGGCTCCCTCCACCGGCACACA 387
Qy	448 gtggcttcacgtctaccagttcttctgtctatcttcaggaaggaagtcctctctcc 507
Db	386 GTGGCTTCATCGTACAGTCTTGTGTATCTTCAGGAAGGAAAGTCATCTCTCC 327
Qy	508 ttcccaagaaacaaactcagagctcttggaatggacagattttcgaacggtttcc 567
Db	326 TTCCCAAGAAACAAAACACTCGAGGCTCTTGAAATGGACAGATTTCTGAACCGTTCC 267
Qy	568 acctgggggaacctgaagcaagcaccagttcatgaccagaactaccaggaactcacaa 627
Db	266 ACCTGGGGGAACCTGAAGCAAGCACCAGTTCATGACCCAGCACTACCGAGACTCACCA 207
Qy	628 ccctccaggtctccagagaaaggccagcgagcccaagcaacaaacccagcgagatag 687
Db	206 CCCFCCAGGCTCCAGAGAAAGGGCCAGCGCCAGCAACCAACCAACCGGAGATAG 147
Qy	688 ctgctgctagatagccgggtttgccatccgggcatgtggccacactccaccaccgac 747
Db	146 CTGCTGCTAGATAGCCGGCTTTCGCATCCGGGCGATGTGGCCACACTCCCAACCGAC 87
Qy	748 gatgtgggtatggaacccctcttgatagacagacccctctttccaaataaaaaaaa 807
Db	86 GATGTGGGTATGGAACCCCTCTCTGGATACAGAACCCCTCTTTTCCAAATAAAAAANA 27
Qy	808 tcatecca 814
Db	26 TCATCCA 20
RESULT	2
LOCUS	AA903899/c
DEFINITION	oe78e07.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1417764 3', mRNA sequence.
ACCESSION	AA903899
VERSION	AA903899.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
	Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 552)
AUTHORS	NCI-CGAP
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1054 Std Error: 0.00

REFERENCE 1 (bases 1 to 440)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407515.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution Information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 1044 Std Error: 0.00  
Seq primer: -40m13 fwd. Et from Amersham  
High quality sequence stop: 440.  
Location/Qualifiers  
1..440  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1417764"  
/clone\_lib="NCI CGAP Lu5"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; 1st strand cDNA was prepared from  
neuroendocrine lung carcinoid, and was then primed with a  
Not I - oligo(dT) primer. Double-stranded cDNA was ligated  
to Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pT73 vector. Library is normalized. Library was  
constructed by Bento Soares and M. Fatima Bonaldo."

FEATURES  
source

1..440  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1417764"  
/clone\_lib="NCI CGAP Lu5"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; 1st strand cDNA was prepared from  
neuroendocrine lung carcinoid, and was then primed with a  
Not I - oligo(dT) primer. Double-stranded cDNA was ligated  
to Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pT73 vector. Library is normalized. Library was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 80 a 93 c 137 g 130 t  
ORIGIN  
Query Match 28.6%; Score 236; DB 39; Length 440;  
Best Local Similarity 99.2%; Pred. No. 5.3e-101;  
Matches 386; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 426 gtcctctcccccgcgcacacagtgctccatcgctaccagttcttctctatctcag 485  
DB 394 gctccctcccccgcgcacacagtgctccatcgctaccagttcttctctatctcag 335  
QY 486 gaaggaaggtctctctctctcccaaggagaaacaaactcgaggtcttggaaaatg 545  
DB 334 GAAGGAAGATCATCTCTCTCTCCCAAGGAAACAAACCTCGAGGCTCTGGAAAATG 275  
QY 546 gacagattctgaacggtttccacgtggcgaaacctgaagcagcaccagttcatgacc 605  
DB 274 GACAGATTTCTGAACGGTTTCCACCTGGCGAACCTGAAGCAAGCACCCAGTTCTATGACC 215  
QY 606 cagaactacaggactcaccacacctccaggctcccgagagaaaggcagcgagcccaag 665  
DB 214 CAGAATACAGGACTCACCAACCTCCAGGCTCCAGAGAAAGGCGCCAGGCCCAAG 155  
QY 666 cacaacaaacaggcgagatagctgctctagatagcggcttttccctccgggcatgt 725  
DB 154 CACAAAACAGCGGAGATAGTGCTGCTGCTAGATAGACCGCTTTGCCATCCGGGATGT 95  
QY 726 ggcacactgccaccacacgaagatgtgggtatggaaacccctcttgatcacagaacccct 785  
DB 94 GGCACACTGCCACCACCGACGATGGGTATGGAACCCCTCTTGATACAGAACCCCT 35  
QY 786 tcttttccaaataaaaaaaatcatcca 814  
DB 34 TCTTTTCCAAATAAAAAAATCATCCA 6

## RESULT 3

F26660 280 bp mRNA EST 13-MAY-1999  
LOCUS HSPD14219 HM3 Homo sapiens cDNA clone s4000058D11, mRNA sequence.  
DEFINITION F26660  
ACCESSION F26660  
VERSION F26660.1 GI:4812286  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 280)  
AUTHORS Lanfranchi, G., Muraro, F., Caldara, F., Pacchioni, B., Pallavicini, A.,  
Pandolfo, D., Toppo, S., Trevisan, S., Scarso, S. and Valle, G.  
TITLE Identification of 4370 expressed sequence tags from a  
3'-end-specific cDNA library of human skeletal muscle by DNA  
sequencing and filter hybridization  
Genome Res. 6 (1), 35-42 (1996)  
96276048  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948109.  
Contact: Valle G.  
CRIBI Biotechnology Centre  
University of Padua  
Via Trieste 75, 35121 Padua, Italy  
ABI Chromatograms and other information are available on WWW at  
<http://group.bio.unipd.it>.  
Location/Qualifiers  
1..280  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="s4000058D11"  
/clone\_lib="HM3"  
/sex="female"  
/tissue\_type="pectoral muscle (after mastectomy)"  
/note="Vector: pCDNAII (Invitrogen); Site 1: BstXI;  
Site 2: NotI; The library was constructed by G.  
Lanfranchi. This library is not subtracted nor normalized.  
The first strand cDNA was primed with a biotinylated  
oligo-dT-NotI primer  
(5'-biotin-AACCGGCTCGAGCGCGCTTTTCTTTTCTTTT-3'). The  
ds cDNA was sonicated and size-selected in the range  
350-550 bp. The 3' specific fragments were selected by  
streptavidin coated magnetic beads, ligated to  
non-palindromic BstXI adapters, NotI digested and  
directionally cloned into BstXI-NotI cut pCDNAII vector."

BASE COUNT  
ORIGIN

Query Match 26.0%; Score 215; DB 50; Length 280;  
Best Local Similarity 99.6%; Pred. No. 4.4e-91;  
Matches 265; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 388 aagggaagattcaaggccaggaggttatcagcctaccaggtctctcccccacgcgcacaca 447  
DB 15 AAGGGAAGATTACAGGCCAGGAGTTATCAGCCTACAGGCTCCTCCACCGGCACACA 74  
QY 448 gtgggttccatcgctaccagttcttctctatcttcaggaaagagtgctatctctctcc 507  
DB 75 GTGGGTTCCATCGCTACCAAGTTCTTTTGTATCTTCAGGAAGAAAGTCATCTCTCC 134  
QY 508 ttcccaagaaacaaacaaactcgagctcttgaaaatgacagattctgaacccctttcc 567  
DB 135 TTCCCAAGAAACAAACAACTCGAGGCTCTTGAAAATGGACAGATTCTGAAACCGTTCC 194  
QY 568 acctggggaacctgaagcaagcaccagttcatgaccagaaactaccaggactcaccaa 627  
DB 195 ACCTGGCGAACCTGAAGCAAGCACCCAGTTCTATGACCACCAACTACCAAGGACTCACCA 254  
QY 628 cctccaggctcccgagagaaaggcc 653  
|||||

1







```

source
1. 423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="s400032A04"
/clone_lib="HM3"
/sex="female"
/tissue_type="pectoral muscle (after mastectomy)"
/notes="vector: pcDNAII (Invitrogen); Site 1: BstXI;
Site 2: NotI; The library is not subtracted by G.
Lanfranchi. This library is not subtracted nor normalized.
The first strand cDNA was primed with a biotinylated
oligo-dT-NotI primer
(5'-biotin-AACCGGCTCGAGCGCGCTTTT-3'). The
ds cDNA was sonicated and size selected in the range
350-550 bp. The 3' specific fragments were selected by
streptavidin coated magnetic beads, ligated to
non-palindromic BstXI adapters, NotI digested and
directionally cloned into BstXI-NotI cut pcDNAII vector."
BASE COUNT 114 a 130 c 103 g 76 t
ORIGIN

Query Match 23.0%; Score 190; DB 50; Length 423;
Best Local Similarity 99.6%; Pred. No. 2.7e-79;
Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 435 ccacggcacacagtggttcctcgcctaccaggtttgttctatcttcaggagaaag 494
|||||
Db 87 CCACGGCACACAGTGGTTCCTCCATCGCTACCAAGTCTTTGTCTATCTTCAGGAAGGAAA 146

Qy 495 gtcatctctctctcccaagaaacaaactcaggtcttctggaatggacagatt 554
|||||
Db 147 GTCATCTCTCTCTCCCAAGAAACAAACTCAGGCTCTGTGAAATGGACAGATT 206

Qy 555 ctgaaccttccacctggcgaaactgaagcaagcaccagttcatcaccagaactac 614
|||||
Db 207 CTGAACCGTTTCCACCTGGCGAACCTGAGCAAGCACCAGTTCATGCCAGAACATAC 266

Qy 615 caggactcaccacacctccaggtcccgagaaagggccagagcccaagcacaacac 674
|||||
Db 267 CAGGACTCACCACCTCCAGGCTCCAGAGAAAGGGCCAGGAGCCCAAGCACAACAAAC 326

Qy 675 c 675
Db 327 c 327

RESULT 12
F31628
LOCUS F31628 258 bp mRNA EST 13-MAY-1999
DEFINITION HSPD22948 HM3 Homo sapiens cDNA clone s4000123B06, mRNA sequence.
ACCESSION F31628
VERSION F31628.1 GI:4817254
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 258)
AUTHORS Lanfranchi, G., Muraro, T., Caldara, F., Pacchioni, B., Pallavicini, A.,
Pandolfo, D., Toppo, S., Trevisan, S., Scarso, S. and Valle, G.
TITLE Identification of 4370 expressed sequence tags from a
3'-end-specific cDNA library of human skeletal muscle by DNA
sequencing and filter hybridization
JOURNAL Genome Res. 6 (1), 35-42 (1996)
MEDLINE 96276048
COMMENT On May 18, 1998 this sequence version replaced gi:3137792.
Contact: Valle G.
CIRIBI Biotechnology Centre
University of Padua
Via Trieste 75, 35121 Padua, Italy
ABI Chromatograms and other information are available on WWW at
http://grup.bio.unipd.it.

FEATURES
source
1. .423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="s400032A04"
/clone_lib="HM3"
/sex="female"
/tissue_type="pectoral muscle (after mastectomy)"
/notes="vector: pcDNAII (Invitrogen); Site 1: BstXI;
Site 2: NotI; The library is not subtracted nor normalized.
The first strand cDNA was primed with a biotinylated
oligo-dT-NotI primer
(5'-biotin-AACCGGCTCGAGCGCGCTTTT-3'). The
ds cDNA was sonicated and size selected in the range
350-550 bp. The 3' specific fragments were selected by
streptavidin coated magnetic beads, ligated to
non-palindromic BstXI adapters, NotI digested and
directionally cloned into BstXI-NotI cut pcDNAII vector."
BASE COUNT 114 a 130 c 103 g 76 t
ORIGIN

Query Match 23.0%; Score 190; DB 50; Length 423;
Best Local Similarity 99.6%; Pred. No. 2.7e-79;
Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 435 ccacggcacacagtggttcctcgcctaccaggtttgttctatcttcaggagaaag 494
|||||
Db 87 CCACGGCACACAGTGGTTCCTCCATCGCTACCAAGTCTTTGTCTATCTTCAGGAAGGAAA 146

Qy 495 gtcatctctctctcccaagaaacaaactcaggtcttctggaatggacagatt 554
|||||
Db 147 GTCATCTCTCTCTCCCAAGAAACAAACTCAGGCTCTGTGAAATGGACAGATT 206

Qy 555 ctgaaccttccacctggcgaaactgaagcaagcaccagttcatcaccagaactac 614
|||||
Db 207 CTGAACCGTTTCCACCTGGCGAACCTGAGCAAGCACCAGTTCATGCCAGAACATAC 266

Qy 615 caggactcaccacacctccaggtcccgagaaagggccagagcccaagcacaacac 674
|||||
Db 267 CAGGACTCACCACCTCCAGGCTCCAGAGAAAGGGCCAGGAGCCCAAGCACAACAAAC 326

Qy 675 c 675
Db 327 c 327

RESULT 12
F31628
LOCUS F31628 258 bp mRNA EST 13-MAY-1999
DEFINITION HSPD22948 HM3 Homo sapiens cDNA clone s4000123B06, mRNA sequence.
ACCESSION F31628
VERSION F31628.1 GI:4817254
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 258)
AUTHORS Lanfranchi, G., Muraro, T., Caldara, F., Pacchioni, B., Pallavicini, A.,
Pandolfo, D., Toppo, S., Trevisan, S., Scarso, S. and Valle, G.
TITLE Identification of 4370 expressed sequence tags from a
3'-end-specific cDNA library of human skeletal muscle by DNA
sequencing and filter hybridization
JOURNAL Genome Res. 6 (1), 35-42 (1996)
MEDLINE 96276048
COMMENT On May 18, 1998 this sequence version replaced gi:3137792.
Contact: Valle G.
CIRIBI Biotechnology Centre
University of Padua
Via Trieste 75, 35121 Padua, Italy
ABI Chromatograms and other information are available on WWW at
http://grup.bio.unipd.it.

FEATURES
source
1. .423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="s400032A04"
/clone_lib="HM3"
/sex="female"
/tissue_type="pectoral muscle (after mastectomy)"
/notes="vector: pcDNAII (Invitrogen); Site 1: BstXI;
Site 2: NotI; The library is not subtracted nor normalized.
The first strand cDNA was primed with a biotinylated
oligo-dT-NotI primer
(5'-biotin-AACCGGCTCGAGCGCGCTTTT-3'). The
ds cDNA was sonicated and size selected in the range
350-550 bp. The 3' specific fragments were selected by
streptavidin coated magnetic beads, ligated to
non-palindromic BstXI adapters, NotI digested and
directionally cloned into BstXI-NotI cut pcDNAII vector."
BASE COUNT 114 a 130 c 103 g 76 t
ORIGIN

Query Match 22.6%; Score 187; DB 50; Length 258;
Best Local Similarity 99.6%; Pred. No. 7.1e-78;
Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 372 ggccggcagcttcccaagaaaggaattcagggccagaggttatcagctaccaggtccc 431
|||||
Db 21 GGCGCGACCTTGAGAAAGGGAAGATTCAGGCGCAGGAGTTATCAGCTACAGGCTCCC 80

Qy 432 tccccaccggcacagtggttccatcgctaccaggtttgttctatcttcaggagaa 491
|||||
Db 81 TCCCGACCGCACAGTGGCTTCATCGCTACCAAGTCTTTGTCTATCTTCAGGAAGGA 140

Qy 492 aagtcattctctctctcccaagaaaggaattcagggccagaggtttgttctatcttcaggagaa 551
|||||
Db 141 AAGGTCATCTCTCTCTCCCAAGAAACAAACTCAGGCTCTTTGGAAAATGGACAGA 200

Qy 552 ttctgaaccttccacctggcgaaactgaagcaagcaccagttcatgaccacaga 609
|||||
Db 201 TTTCTGACCGCTTCCACCTGGCGAACCTGAGCAAGCACCAGTTCATGACCAGA 258

RESULT 13
F20527
LOCUS F20527 395 bp mRNA EST 13-MAY-1999
DEFINITION HSPD04880 HM3 Homo sapiens cDNA clone NOTAVAIL04880, mRNA sequence.
ACCESSION F20527
VERSION F20527.1 GI:2059703
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 395)
AUTHORS Lanfranchi, G., Muraro, T., Caldara, F., Pacchioni, B., Pallavicini, A.,
Pandolfo, D., Toppo, S., Trevisan, S., Scarso, S. and Valle, G.
TITLE Identification of 4370 expressed sequence tags from a
3'-end-specific cDNA library of human skeletal muscle by DNA
sequencing and filter hybridization
JOURNAL Genome Res. 6 (1), 35-42 (1996)
MEDLINE 96276048
COMMENT On Apr 14, 1993 this sequence version replaced gi:635805.
Contact: Valle G.
CIRIBI Biotechnology Centre
University of Padua
Via Trieste 75, 35121 Padua, Italy
ABI Chromatograms and other information are available on WWW at
http://grup.bio.unipd.it
POLYA=NO.
Location/Qualifiers
1. .395

```



```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NOTAVAIL04880"
/clone_lib="HM3"
/sex="female"
/tissue_type="pectoral muscle (after mastectomy)"
/notes="Vector: pcDNAII (Invitrogen); Site 1: BstXI; Site 2: NotI; The library is not subtracted nor normalized. The first strand cDNA was primed with a biotinylated oligo-dT-NotI primer (5'-biotin-AACCGGCTCGAGCGCGCTTTTTTTTTTTT-3'). The ds cDNA was sonicated and size-selected in the range 350-550 bp. The 3' specific fragments were selected by streptavidin coated magnetic beads, ligated to non-palindromic BstXI adapters, NotI digested and directionally cloned into BstXI-NotI cut pcDNAII vector."
BASE COUNT      109 a      113 c      98 g      71 t      4 others
ORIGIN

```

```

Query Match      22.6%; Score 187; DB 33; Length 395;
Best Local Similarity 100.0%; Pred. No. 6.9e-78;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 544 tggacagattttgaaacggtttccaccctggcgaaacctgaagcaagcaccagttcatga 603
Db 209 TGGACAGATTTCGTGACCGTTTCCACCTGGGCGAACCCTGAAGCAAGCACCCAGTTCATGA 268

Qy 604 cccgaactaccagactcaccacccctccaggtcccgagaaagggccagcgagccca 663
Db 269 CCCGAACCTACCAGGACACCAACCCCTCCAGGCTCCCGAGAAAGGGCCAGCGAGCCCA 328

Qy 664 agcaaaaaaacaggcgagatagctgctgtagatagcggtttgcccattccgggcat 723
Db 329 AGCAAAAAACAGCGCGAGATAGTGCCTGCTAGATAGCGGCTTGCCATCCGGGCAT 388

Qy 724 gtggcca 730
Db 389 GTGGCCA 395

```

```

RESULT 14
F24489
LOCUS      F24489      334 bp      mRNA      EST      13-MAY-1999
DEFINITION HSPD10829 HM3 Homo sapiens cDNA clone s4000013A01, mRNA sequence.
ACCESSION  F24489
VERSION     F24489.1 GI:4810115
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 334)
AUTHORS   Lanfranchi,G., Muraro,T., Caldara,F., Pacchioni,B., Pallavicini,A.,
Pandolfo,D., Toppo,S., Trevisan,S., Scarso,S. and Valle,G.
TITLE     Identification of 4370 expressed sequence tags from a
3'-end-specific cDNA library of human skeletal muscle by DNA
sequencing and filter hybridization
JOURNAL   Genome Res. 6 (1), 35-42 (1996)
MEDLINE   96276048
COMMENT   On Jun 5, 1998 this sequence version replaced gi:3188833.
Contact: Valle G.
CRIBI Biotechnology Centre
University of Padua
Via Trieste 75, 35121 Padua, Italy
ABI Chromatograms and other information are available on WWW at
http://group.bio.unipd.it.
Location/Qualifiers
1. .334
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="s4000013A01"

```

## FEATURES

source

```

/clone_lib="HM3"
/sex="female"
/tissue_type="pectoral muscle (after mastectomy)"
/notes="Vector: pcDNAII (Invitrogen); Site 1: BstXI; Site 2: NotI; The library is not subtracted nor normalized. The first strand cDNA was primed with a biotinylated oligo-dT-NotI primer (5'-biotin-AACCGGCTCGAGCGCGCTTTTTTTTTTTT-3'). The ds cDNA was sonicated and size-selected in the range 350-550 bp. The 3' specific fragments were selected by streptavidin coated magnetic beads, ligated to non-palindromic BstXI adapters, NotI digested and directionally cloned into BstXI-NotI cut pcDNAII vector."
BASE COUNT      97 a      103 c      76 g      58 t
ORIGIN

```

```

Query Match      22.5%; Score 186; DB 50; Length 334;
Best Local Similarity 99.3%; Pred. No. 2.1e-77;
Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 388 aaggaaatttcaggccagcaggttatcagctaccaggtccctcccccgcgcacaca 447
Db 44  AAGGAAGATTTCAGGCGCCAGGAGTTATCAGCCTACCGGCTCCCTCCCGGCACACA 103

Qy 448 gtggcttcacgtcgtaccagttcttctgtctatcttcagaaaggaagtcattctctcc 507
Db 104 GTGGCTTCACGTCTACCAGTTCITTTGCTATCTTCAGGAAGGAAGTCACTCTCTCC 163

Qy 508 ttcccaaggaaaaaaactcgaggtcttggaaaatggacagatttgcacgtttcc 567
Db 164 TTCCCAAGGAACAAAACCTTCGAGGCTCTTGAGAAATGACAGATTTCTGAACCGTTTCC 223

Qy 568 acctggggcgaacctgaagcaagcagccagtcagccagaaactaccagactcaccaa 627
Db 224 ACCTGGGGCGAACCTGAAGCAAGCACCAGTTCATGACCCAGAACTTACCAGGACTCACCAA 283

Qy 628 ccctcaggtctcccgagagaagggcgagcgagcccaagcacaacaaacc 675
Db 284 CCCTCAGGCTCCCGAGAGAAAGGGCGCAGCGAGCCCAAGCACAACAAACC 331

```

```

RESULT 15
W37255
LOCUS      W37255      452 bp      mRNA      EST      10-OCT-1996
DEFINITION zC09f01.t1 Soares.Parathyroid tumor.NbHPA Homo sapiens cDNA clone
IMAGE:321817 5' similar to SW:PBPL_RAT P31044 23 KD
MORPHINE-BINDING PROTEIN ;, mRNA sequence.
ACCESSION  W37255
VERSION     W37255.1 GI:1318849
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 452)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Woldmann,P. and
Wilson,R.
TITLE     The WashU-Merck EST Project
JOURNAL   Unpublished (1995)
COMMENT   On May 8, 1995 this sequence version replaced gi:801082.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

```

```

Insert Length: 569 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 434.
Location/Qualifiers
1..452
/organism="Homo sapiens"
/db_xref="GDB:1259475"
/db_xref="taxon:9606"
/clone="IMAGE:321817"
/clone_lib="Soares.parathyroid.tumor_NbHPA"
/tissue_type="parathyroid.tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/Note="Organ: parathyroid gland; Vector: pT7T3D
(Pharmacia) with a modified polylinker; Site.1: Not I;
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer
[5',
TGTATCAATCTGAAGTGGAGCGCGCCACCAATTTTTTTTTTTTTTTTTTTT
T-3', double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT7T3
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."
BASE COUNT 125 a 137 c 108 g 80 t 2 others
ORIGIN

```

```

Query Match 22.5%; Score 186; DB 26; Length 452;
Best Local Similarity 99.6%; Pred. No. 2e-77;
Matches 236; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 428 tccctccaccgacacagtggttcacatgcctaccagttcttcttatcttcagga 487
Db 101 TCCTCCCGCCGACACAGTGGCTTCCATCGCTACCACTTCTTGTCTATCTTCAGGA 160
QY 488 aggaagggtcatctctctcccaaggaacaaactcgaggctcttgaaaaatgga 547
Db 161 AGGAAAGTCACTCTCTCTCTCCCAAGGAACAAACTCGAGGCTCTTGGAAATGGA 220
QY 548 cagattctgaacggtttccacactggcgcaacctgaagcaagcaccacagttcatgaccca 607
Db 221 CAGATTTCTGAACCGTTTCCACCTGGCGCAACCTGAAGCAAGCACCCAGTTCATGACCCA 280
QY 608 gaactaccaggactaccacacctccaggctccagagaaaaggccagagcccaa 664
Db 281 GAATACCAAGGACTACCAACCTCCAGGCTCCCAAGAAAGGGCCAGCGACCCAA 337

```

Search completed: May 1, 2000, 14:45:01  
Job time: 14021 sec